FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

HealthTalks - Improving Health Communication and Personal Information Management

João M. Monteiro

Mestrado Integrado em Engenharia Informática e Computação

Supervisor: Carla Teixeira Lopes

July 10, 2017

The present thesis has been developed within the project "NORTE-01-0145-FEDER-000016", financed by the North Portugal Regional Operational Programme (NORTE 2020), under the PORTUGAL 2020 Partnership Agreement, and through the European Regional Development Fund (ERDF).

HealthTalks - Improving Health Communication and Personal Information Management

João M. Monteiro

Mestrado Integrado em Engenharia Informática e Computação

Approved in oral examination by the committee:

Chair: João Correia Lopes

External Examiner: Ana Margarida Almeida

Supervisor: Carla Teixeira Lopes

July 10, 2017

Abstract

A patient's health literacy, which is their ability to obtain and use health information, has a direct impact on their health, but half of the Portuguese population has problematic or inadequate levels of health literacy, whereas in the USA more than a third of the population has basic or below basic levels of health literacy. An individual's wellbeing is also affected by their relationship with their physician and their personal health information management proficiency. When it comes to the communication between a patient and their doctor, it works best when it is focused on decision-making, but when that does not happen patients tend to distrust and double-check their doctors, and the latter are not always receptive to such behaviour. Patients may also have trouble managing their personal health information due to methodology failures or the lack of means to do so adequately.

Technological solutions can have a positive impact in some of those aspects due to factors such as the ubiquity of mobile devices nowadays. In this project we proposed and developed a mobile app to help correct the faults in the patients' healthcare. HealthTalks empowers the patients with tools that may ease their day-to-day health tasks and self-care ability. It does so by enabling a better personal health information management and promoting a more straightforward communication between the patient and the physician, while providing more information about the health topics they discuss. This is accomplished by recording the audio of a medical appointment, transcribing its dialogue, giving more information about used medical concepts, and allowing all of that information to be easily managed by the patient.

To characterise and better understand the app's target population, we conducted a survey online and in the major hospital of the North of Portugal. This survey focused on the note-taking behaviours of patients regarding their personal health information, the relationship they have with their doctor, and the conduct that both display during a medical appointment. We also inquired about the app's features in order to gauge which components they perceived as more useful. The results show that the majority of patients do not take notes of their health information, and that the biggest obstacle to patients understanding their doctors is the latter's use of medical and technical concepts.

Even though the app is meant for the general adult Portuguese population, extensive usability tests with full demographic representation were not feasible, so we conducted those tests exclusively with elderly users. This is a target group that is expected to benefit greatly from this solution, while at the same time assumed to be less receptive to it. These tests allowed us to define and rethink some aspects related to HealthTalks' interface and user experience, facilitating its adoption by a wider range of users.

There were several options of speech recognition software to implement, which were narrowed to two after comparing their characteristics: Google Cloud Speech API and Bing Speech API. Those options were tested with generalist and medical texts so that we could make an informed final decision on which to use. Our choice was Google Cloud Speech API, since it consistently showed better results than the latter.

After its development, the app's performance was also evaluated to understand its applicability in a Portuguese context. That evaluation was mainly focused on the quality of the transcriptions, since the definition of terms depends on it being accurate in order to correctly identify medical terminology. The result was a 12% word error rate when transcribing a medical text. There was an option to send specific medical phrases to the speech recogniser in order to improve its accuracy; however, our tests showed that doing so was not necessary since the vast majority of medical concepts were correctly transcribed without it.

In the future, our goal is to simulate appointments to better reproduce their real conditions, add more features to the app, and conduct extensive interviews with physicians so as to better understand their perspective on the problems previously mentioned and HealthTalks' approach to tackling them.

Resumo

A literacia em saúde de um paciente, ou a sua capacidade de obter e utilizar informação de saúde, tem um impacto direto na sua saúde; contudo, metade da população portuguesa tem níveis problemáticos ou inadequados de literacia em saúde, enquanto nos E.U.A. mais de um terço da população tem um nível básico ou inferior de literacia em saúde. O bem-estar de um indivíduo também é influenciado pela sua relação com o médico e a sua capacidade de gestão de informação pessoal de saúde. Relativamente à comunicação entre o paciente e o médico, um funcionamento ótimo acontece quando se foca na tomada de decisão, mas quando isso não acontece os pacientes tendem a duvidar dos seus médicos e confirmar as suas afirmações, comportamento esse que muitos médicos não aceitam. Os pacientes podem também ter dificuldade em gerir a sua informação pessoal de saúde devido a falhas de metodologia ou a falta de meios para o fazer adequadamente.

Algumas soluções tecnológicas podem ter um impacto positivo em alguns desses aspetos devido a fatores como a ubiquidade de dispositivos móveis hoje em dia. Neste projeto propusemos e implementámos uma aplicação para dispositivos móveis que ajuda a corrigir os problemas nos cuidados de saúde dos pacientes. A HealthTalks fornece aos pacientes ferramentas que poderão melhorar as suas condições de saúde. Fá-lo através de um sistema de gestão de informação pessoal de saúde, promovendo uma comunicação mais clara entre médico e paciente, e fornecendo informações sobre os tópicos que discutem. Estes objetivos são alcançados através da gravação do áudio de consultas médicas, a transcrição destas, e a disponibilização de informação sobre termos médicos, sendo toda essa informação facilmente gerível pelo paciente.

Para melhor caracterizar e compreender a população-alvo da aplicação, distribuímos um inquérito no maior hospital do norte de Portugal e online. Este inquérito focou-se nos comportamentos de anotação de informações de saúde dos pacientes, na relação que têm com o seu médico, e nos comportamentos de ambos durante uma consulta. Também os questionámos relativamente às funcionalidades da aplicação, de modo a perceber melhor que componentes eram consideradas mais úteis. Os resultados mostram que a maioria dos pacientes não toma nota das suas informações pessoais de saúde, e que o maior obstáculo na compreensão dos médicos é o uso que estes fazem de termos médicos e técnicos.

Apesar de a aplicação ser destinada ao público adulto português, seria impraticável testar a sua usabilidade em todas as faixas demográficas. Assim, escolheu-se testá-la exclusivamente com idosos, uma vez que estes são o público-alvo que se espera ser mais beneficiado, e simultaneamente aquele que poderá ser menos recetivo a esta solução. Estes testes permitiram-nos definir e repensar alguns aspetos da interface e usabilidade da HealthTalks, potenciando a sua adoção por um maior número de utilizadores.

Várias opções de software de reconhecimento de voz foram consideradas, mas duas foram selecionadas depois de comparar as suas características: Google Cloud Speech API e Bing Speech API. Estas foram testadas com um texto generalista e outro médico de forma a que pudesse ser tomada uma decisão informada. A nossa escolha recaiu na primeira opção, visto que mostrou resultados consistentemente melhores.

O desempenho final da aplicação foi avaliado para estudar a sua aplicabilidade no cenário português. Esta avaliação focou-se sobretudo na qualidade da transcrição, uma vez que esta é necessária para a correta identificação de terminologia médica. O resultado foi uma taxa de erro de 12% ao transcrever um texto médico. Havia a possibilidade de enviar conceitos médicos para o software de reconhecimento de voz de modo a melhorar a sua precisão; no entanto, os nossos testes demonstraram que tal era desnecessário uma vez que a grande maioria dos termos médicos era corretamente transcrita sem essa funcionalidade.

No futuro pretendemos simular consultas médicas para reproduzir as suas condições reais, acrescentar mais funcionalidades à aplicação, e conduzir entrevistas extensivas com profissionais de saúde para perceber quais são as suas perspetivas relativamente aos problemas na prestação de cuidados de saúde já mencionados e à forma como a HealthTalks os aborda.

Acknowledgements

First and foremost, I would like to thank my supervisor, Professor Carla Teixeira Lopes, for her unwavering encouragement, support, and guidance over the last few months. She trusted me with this project, and I hope my work and suggestions have enriched her initial ideas.

I would also like to thank my family and friends for being a continued presence throughout these months, helping and inspiring me every day. A special thanks goes to my parents, to whom I owe everything I have accomplished; my sisters for their continued presence; Flávio for being a source of strength and motivation when I faced adversities; and Nelson, who is now surely inadvertently more knowledgeable of computer science and statistics after being there for the whole journey.

Thanks also to Dr. Dagmara Paiva, whose help was invaluable in developing new questions for the questionnaire and adapting its vocabulary to be more accessible.

To Dr. João Amaro, our connection link in the São João Hospital Centre (CHSJ); CHSJ's Health Ethics Committee; Dr. Xavier Barreto, director of CHSJ's ambulatory centre; and Tiago Nércio; my thanks for allowing and helping us distribute the printed version of the questionnaire in CHSJ.

A big thanks also to the Integrated Masters in Informatics and Computing Engineering course, and in particular Professor João Pascoal Faria, for subsidising the printing of the questionnaires and the poster presented in DSIE. To D. Lina and D. Sandra, my thanks for your help, as well as all of the staff's patience, during the printing and stapling of all those copies during all those afternoons.

Professor A. Parasuraman, Mr. Charles Colby, and Rockbridge Associates, Inc. have my gratitude for allowing me to use their world-renowned Technology Readiness Index at no cost in my research.

I would like to thank the Doctoral Symposium on Informatics Engineering and its organisers, for allowing our project to be shown and discussed.

Finally, my very grateful appreciation for all the friends, family members, colleagues, staff, patients, and strangers who answered the questionnaire. A special thanks to the five senior citizens who, despite some initial hesitation, were kind enough to help me test HealthTalks' usability.

João M. Monteiro

"Developments in medical technology have long been confined to procedural or pharmaceutical advances, while neglecting a most basic and essential component of medicine: patient information management."

John Doolittle

Contents

List of Figures

List of Tables

LIST OF TABLES

Abbreviations

API	Application Programming Interface
APK	Android Package
App	Mobile software application
ASR	Automatic Speech Recognition
BSA	Bing Speech API
CES	Comissão de Ética para a Saúde (Health Ethics Committee)
CHSJ	Centro Hospitalar São João (São João Hospital Centre)
DB	Database
DSIE	Doctoral Symposium on Informatics Engineering
EU	European Union
FEUP	Faculdade de Engenharia da Universidade do Porto (Engineering Faculty of the University of Porto)
GCSA	Google Cloud Speech API
HCI	Human-computer interaction
HMM	Hidden Markov Models
IDE	Integrated Development Environment
JSON	JavaScript Object Notation
MIEIC	Mestrado Integrado em Engenharia Informática e Computação (Integrated Masters in Informatics and Computing Engineering)
MWA	MediaWiki API
ODPHP	Office of Disease Prevention and Health Promotion
PHIM	Personal Health Information Management
PIM	Personal Information Management
REST	Representational State Transfer
RPC	Remote Procedure Call
SDK	Software Development Kit
SNS	Serviço Nacional de Saúde (National Health Service)
SQL	Structured Query Language
STT	Speech-to-text
TR	Technology Readiness
TRI	Technology Readiness Index
UC	Use case
UI	User Interface
UP	University of Porto
US	User story
USA	United States of America
UX	User experience
WER	Word Error Rate
WHO	World Health Organization
XML	Extensible Markup Language

Chapter 1

Introduction

Every individual has the right to proper healthcare, which is the responsibility not only of the healthcare institutions and professionals, but also of oneself. People are expected to take care of themselves and others, but not everyone has the necessary skills to understand what their physicians tell them and to make proper decisions. The goal of this thesis was to develop an application that: (1) helps people with lower levels of health literacy better understand medical speech and (2) provides mechanisms to assist personal health information management.

1.1 Context

First of all, it is important to understand that an individual's health literacy affects their ability to understand health information and act on it to maintain or improve their quality of life [SVF+12]. This means that, without an appropriate level of health literacy, a person might not be able to efficiently seek for health information or manage their personal health information correctly. This, in turn, may result in the individual not taking good care of their own and their family's health.

Health literacy levels are not universally high. In Portugal, for example, only half of the population had sufficient or excellent general health literacy levels in 2014 [EÁM15]. Physicians might help these users through simpler messages but the doctor-patient relationship is not always optimal, leading sometimes to miscommunication and lack of trust to confide crucial information [PR00].

An alternative would be using health technologies to complement the patient's self-care skills. Those can range from health information systems that try to contextualise the patient's health concerns, to personal health information management systems that can be used to better organise and retrieve information concerning someone's health conditions.

Even though the use of technology might not always result in better information management, it can do so if kept moderately simple [Jon07]. It is also believed that it can be an informing tool both by exposing data in simple and engaging ways [FF03], as well as by fostering learning through assimilation [LGF+15].

1.2 Motivation and Goals

There are several technologies available to solve the problems outlined in Section 1.1, but they tend to have a narrow focus, specialising in niches or particular health fields. Few are adapted to more realistic and holistic scenarios, such as when a user handles different types of health information, from calories and weight, to blood analysis results and medical treatments. Furthermore, the functionalities are often restrictive, having a lower value for the user due to the lack of customisation and limited potential for innovative uses.

HealthTalks is our solution to tackle those issues, hopefully resulting in better healthcare for everyone, but especially the individuals with lower health literacy. Its premise relies on the recording of appointments through a mobile app. The resulting audio files are stored in the app and incorporated in an appointment page. Each page is linked to one recording and displays that recording's transcription when solicited. It also provides several customisable fields such as its title, the patient involved, the doctor who gave the appointment, their medical speciality, and where it happened. The user can also mark an appointment as a favourite, in case it is important. All of those fields, in tandem with the date of the recording, can be used to organise and search for a specific appointment at a later date. All of these functionalities create a robust personal health information management system while simultaneously requiring little effort from the user.

The app's focus on health information also helps to inform the patient of their conditions. This was achieved by finding the medical terms in each appointment's transcription and providing definitions for each, in case the user wants to learn more about them.

By tackling both health information seeking and personal health information management, keeping the app generic so that it can be used in different healthcare fields, supplying optional definitions of medical terms, and providing flexibility to the users through customisability, we hope to contribute to improved self-care practices.

1.3 Contributions

This work, specifically the details involving the HealthTalks app, its motivations, and the project's goals, has been published in the proceedings of the 12th Doctoral Symposium in Informatics Engineering (DSIE). A poster with content from that article was also presented in that same symposium and was rewarded with the accolade of best poster on January 31st 2017 at the Faculty of Engineering of the University of Porto.

1.4 Report Structure

This dissertation is divided in eight major chapters. In Chapter 2, we focus on health literacy: what it is, why it matters, what are the levels of health literacy in the population, and what can be done to improve those levels.

Chapter 3 focuses on the acquisition and management of health information by the patient. We start by analysing their relationship with the physician, while explaining its relevance and how one could improve it. Then we tackle the health information seeking process, which is about how patients retrieve health information and how efficient they are at doing it. At the end of the chapter we explore the concept of personal information management and, more specifically, personal health information management. We debate on what it entails exactly, while explaining its benefits and shortcomings.

Chapter 4 focuses on technological solutions applied to healthcare. We start by giving an introduction to speech recognition, from its history to the basic workings behind it, and progress to study its applicability in healthcare. In that section, we also highlight five specific mobile applications that we perceive as having a particular significance due to their uniqueness or functionalities.

In Chapter 5 we study the Porto population in regards to their note-taking habits of health information, their relationship with their doctors, and their interest in a solution such as what we propose with HealthTalks. That study was done through the distribution of questionnaires, both online and in São João Hospital Centre, and involved more than 1200 respondents.

Chapter 6 entails the software requirements for HealthTalks, including product requirements, personas, and user stories. Those aspects enabled us to have a better perspective on the solution, which means that we started working on it already aware of its goals and features.

In Chapter 7 we introduce HealthTalks and its components. We start by giving a broad system description, and then detail its architecture, the third-party software used, how we developed its user experience, and what limitations it faces.

Chapter 8 encompasses all of the tests we did over the development of this project. These include a comparison between two speech recognition solutions, one user experience assessment that includes a usability test of the app, and a final speech recognition test in which we evaluate the performance of the chosen speech recognition solution in Portuguese.

Finally, in Chapter 9 we make a brief summary of all the work we have developed over the last months and what we plan to do in the future.

Chapter 2

Health Literacy

Health literacy is a term introduced in the 1970s that has been increasingly valued by governments and researchers worldwide, as evidenced by the creation of more programs and systems to improve it, as well as more research and studies on its sources and consequences [SVF+12]. In this chapter we explain what it entails and why it matters.

2.1 Definition

Different definitions for health literacy have been proposed over the years. Hur et al. [HLS15] define it as an understanding of basic medical information, but we can go farther and maintain that it also entails the ability to seek, process, and interpret basic medical information in order to get proper treatment [KGJ+06]. Sørensen et al. [SVF+12] corroborate that concept by defending that it empowers an individual to think and act upon their healthcare in order to maintain or improve their quality of life. From that statement one can deduce that it is an important skill regardless of an individual being ill, at risk, or just wanting to stay healthy.

The concept of health literacy has shifted in recent years, with some studies claiming that it is also affected by factors external to the individual. In fact, Heijmans et al. [HUR+15] emphasise that an individual's social support system, such as their family and community, and their context (e.g. the healthcare system in their country), are also important in shaping their health literacy.

Osborne [Osb13] refers yet another component: the ability to convey a health-related message. In her view, the communication skills of a person contribute to their own and other's health literacy, which means that one's level of health literacy might affect the message they convey and, hence, the ease of understanding by others. Health knowledge or proficiency is sometimes said to be itself yet another component of health literacy, in addition to the perception of new information already mentioned [Bak06].

It can be therefore inferred that health literacy is not confined to a simple understanding of terms and their meanings, but rather a complex web of interacting components such as knowledge, interpretation, critical thinking, communication, and decision-making, as well as the individual's social network and context. To add to that, other factors such as a person's attitude and willingness to take care of their health may influence their health literacy, even if not inherently part of its definition [HUR+15]. A conceptual model of the different elements often associated with health literacy can be seen in Figure 2.1.

The fact that so many distinct interpretations of the concept exist makes it difficult to properly evaluate and compare health literacy levels measured in different studies, since they may rely on different definitions. Not only that, but if a definition of health literacy includes the context of the subject, then measuring it becomes a much more demanding task, since it cannot be assessed by looking exclusively at the subject's personal data [Bak06].

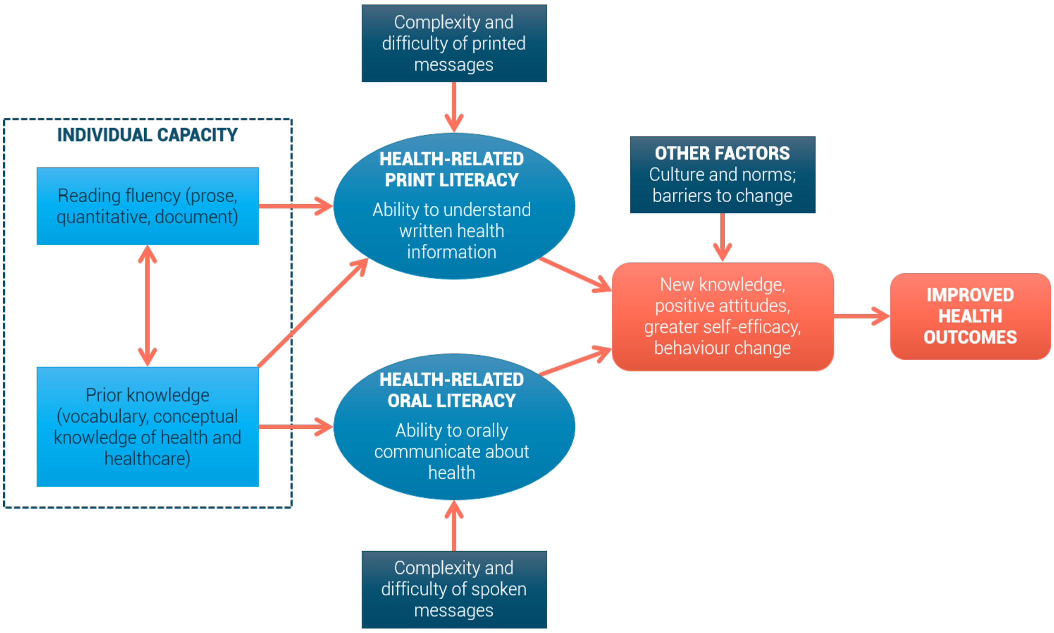


Figure 2.1: "Conceptual model of the relationship between individual capacities, health-related print and oral literacy, and health outcomes." [Bak06] [Adapted]

2.2 Importance

The level of someone's health literacy influences their health behaviours and health conditions [HLS15]. If someone has lower levels of health literacy, they are less likely to understand the importance of preventing illnesses by eating healthily, exercising, and caring for themselves, which may lead to health problems such as obesity, diabetes and heart disease [Car06]. Moreover, they are less likely to get preventive care and more likely to have treatment delays, which results in them going to emergency rooms with more severe symptoms [HLS15]. When getting professional care, individuals with low health literacy are also more vulnerable to misinterpreting their health professional [Car06] and medication instructions [HLS15], which in turn may result in inadequate treatments and self-care.

In summary, higher levels of health literacy have been linked to improved health conditions, reduced costs of healthcare, and reduced use of healthcare services [HUR+15; HLS15; Off16]. By the same token, lower health literacy standings are associated with less knowledge of illnesses and inadequacy in negotiating the healthcare system [JBO10]. In 2006, Carmona [Car06] referred to the unsatisfactory level of health literacy in the United States of America (USA) as a crisis and a national priority due to its impact in every citizen's lives. This has also been referred to as the "health literacy epidemic" by Sørensen et al. [SVF+12].

2.3 Current Situation

Health literacy levels vary according to several demographic variables. Older adults, racial and ethnic minorities, less educated people, poorer people, non-native speakers, and people with weakened health are the most likely to have lower levels of health literacy in the USA [Off16]. But there is evidence of similar statistics in Portugal, being that older people and people with lower levels of education manifest lower levels of health literacy as well [EÁM15].

As can be seen in Figure 2.2, more than a third of Americans had "basic" or "below basic" levels of health literacy in 2003, with the majority of the country having at least an intermediate level. A European Union (EU) study also divides literacy in 4 proficiency groups, from inadequate to excellent. In those statistics, shown in Figure 2.3, Portugal is in 6th place among the 9 countries considered, with almost half of its population presenting "problematic" or "inadequate" levels of health literacy, slightly higher than EU's average of 47.6%. By contrast, the Netherlands has 28.7% of its population with those same levels of health literacy, placing it at the top of the list with the highest levels of health literacy in the considered sample. However, it is important to note that a direct comparison between the USA and EU studies cannot be made, firstly because their nomenclature for each health literacy level is different, and secondly due to the reasons explained in Section 2.1 related to the absence of a universal definition for the term "health literacy".

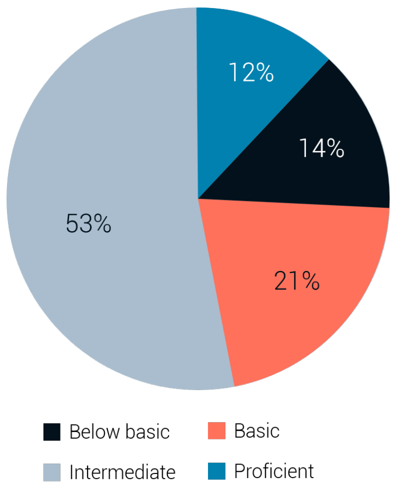


Figure 2.2: Levels of health literacy in the USA in 2003. [Off08] [Adapted]

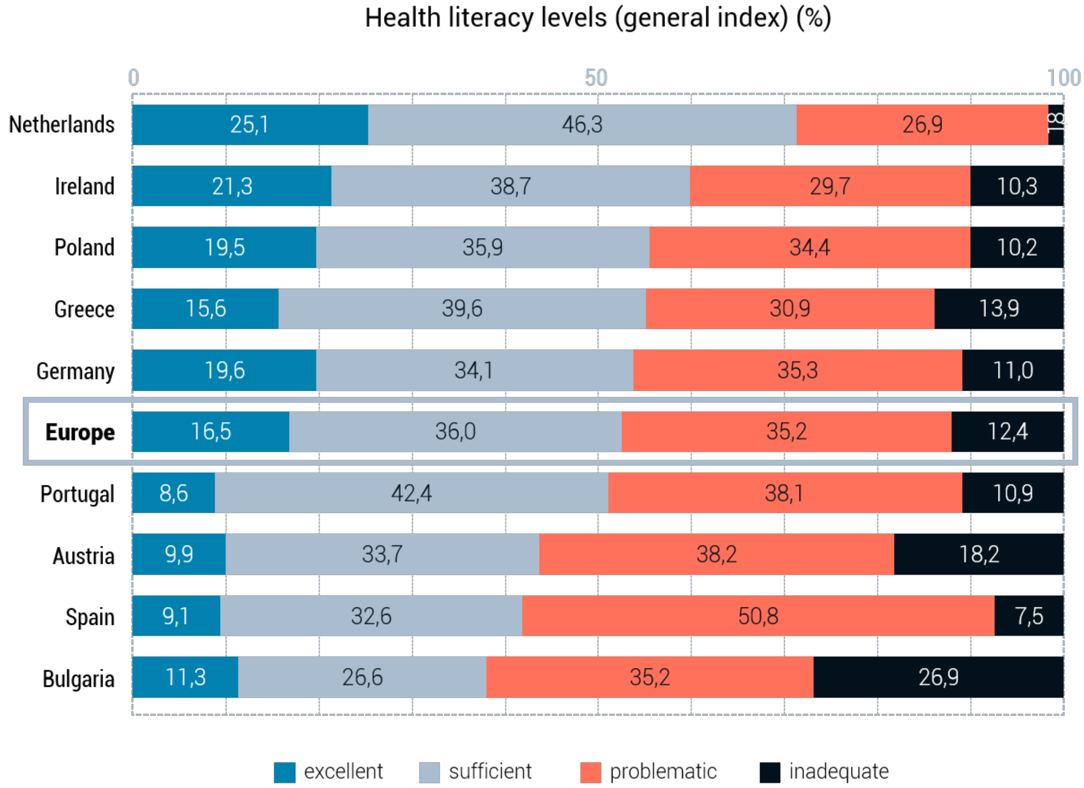


Figure 2.3: Levels of general health literacy in Portugal in 2014, in other select EU countries in 2011, and the European average, highlighted. [EÁM15] [Adapted]

2.3.1 Measures to Improve Health Literacy

The American Medical Association established a list of actions that should be taken in order to improve health literacy levels: evaluate current health literacy levels, improve communication with low-literacy patients, study the consequences of a low level of health literacy, and investigate the causality between a person's poor health literacy and their health condition [SVF+12]. These approaches would give a more in-depth knowledge and understanding of health literacy and its ramifications, which in turn might give clues on how to tackle it on a larger scale.

In Portugal, both the 2011-2016 and the 2016-2020 National Health Plans mention health literacy, the latter promoting it at both national and regional levels. The strategies outlined propose getting the citizens engaged in health institutions and systems in order to improve their health literacy. On top of that, the Portuguese Department of Health Promotion and Prevention of Non-transmittable Diseases divulges prevention and control methods for chronic diseases, and cooperates with the Portuguese National Health System in a collaborative research lab that regularly publishes information regarding disease prevention and health promotion. The Calouste Gulbenkian Foundation also tackles these issues by having a fund specifically for researching and promoting health literacy, targeting particular segments of the population (such as low-education, elders, or prisoners) and health conditions (like diabetes type 2 and obesity in adolescence) [HUR+15].

"Literacia para a Saúde" is a platform maintained by the Lusophone Network to Promote Health Literacy, a result of the cooperation between Portugal and Brazil, to gauge the adequacy of existing models of evaluating and promoting health literacy in the context of these two countries [Red17].

2.4 Prospective Developments

Even though it affects mostly the patients, it is the public health professionals who are responsible for promoting health literacy, according to the Office of Disease Prevention and Health Promotion (ODPHP) [Off16]. They also add that educators should reach out to adults with limited literacy skills, while never neglecting accessibility in health information and services.

The ODPHP [Off08] also defends the creation of guidelines for information access and design, so that that the term "accessible health information" can be defined and adhered to, making the information universally available. The health information materials should be regularly tested with consumers and designed in such a way that their health literacy does not substantially affect their understanding of the contents. The ODPHP also highlights the importance of increasing the communication skill requirements both in secondary education and graduate schools in order to improve the communication between health consumers and professionals.

There are several different types of technologies with the purpose of helping users change their behaviour, make decisions or learn. Their accessibility, manifested in ways such as the simplicity of making an appointment or taking the appropriate time to answer e-mails or phone calls, is an important factor for their success [HLS15]. Moreover, Hur et al. [HLS15] also argue that comprehensibility (through the use of low level English, for example) is more important for explicit knowledge learning than specific technical features.

There is a stark difference in the use of health literacy technologies between young and old users. While the former respond better to engaging technologies (such as online gamification experiences, videogames, and other multimedia), the latter rely on diverse sources of information and tend to favour smartphones and iPads due to their versatility to do other things (such as communicating with healthcare providers, discussing issues, or being entertained). One of the most important aspects of health literacy technologies for this age segment is its ease of use. [HLS15].

2.5 Summary

Even though health literacy is not a recent concept, its meaning and nuances have evolved over time. At the moment it is widely believed that it is a skill that involves the ability of an individual to acquire and understand health information, as well as putting it to proper use on themselves and their communities. However, as more attention is directed toward this area, new studies may expand the concept even further.

The reality of health literacy in Portugal may not be very encouraging, since almost half of the population has problematic or inadequate levels of health literacy, but several organisations, both governmental and independent, have proposed measures and are developing studies in order to better understand and improve health literacy. The same is happening in the USA, though their health literacy levels seem to be higher than in most EU countries.

Health literacy technologies should be adapted to their target audience, especially in regards to age groups. While younger users opt for more involving technologies, the elderly give more importance to the versatility and communication capability of mobile devices.

Chapter 3

Patients and Health Information

Health information refers to all the information pertaining to an individual's fitness and well-being that enables them to live a healthy life. It can be acquired in many ways, from a doctor in a medical appointment, to online search, or specialised media. In this chapter we will analyse the concepts behind health information, its relevance, what may hinder its access, and what means exist to help people finding and keeping it.

3.1 Doctor-Patient Relationship

The doctor-patient relationship has been changing since the second half of the twentieth century from a unilateral exchange in which the physician is the sole contributor, to a more reciprocal, decision-making, or relationship-centred exchange [LB14; Aro03]. This emphasis on the health consumer changes the dynamics of the relationship between them and their health professionals. The latter may be unprepared to deal with information found by the health consumers, leading to stressful and frustrating medical appointments where they have to correct their patients. The second-guessing and questioning of their practitioners may also be seen as a reflection of the patients' diminished trust in them [CH01]. Arora [Aro03], however, advocates that better informed patients will have better health outcomes, use healthcare resources more appropriately, and have better relationships with their health professionals. Fox and Fallows [FF03] endorse both of those theories, asserting that patients who use the Internet to seek information report different — though not always better — relationships with their physicians, depending on the receptivity of the latter to Internet search.

Several studies have shown that a caring, compassionate, respectful, and trusting physician-patient relationship helps cancer patients adjust better to their situation. Its importance may even lead doctors to acquiesce to inappropriate requests from the patients in order not to weaken their bond [Aro03].

3.1.1 Trust

In the realm of physician-patient relationships, trust can have different meanings. Some argue that it refers to the patients' assumptions on how their clinician will behave, whereas others see it as the patient's vote of confidence on the doctor's good intentions. Some people describe their trust as the result of the physician's competence, compassion, privacy, confidentiality, reliability, dependability, honesty, and communication [PR00]. A trusting relationship is beneficial to the patient, since it increases their prospect of treatment compliance [PR00].

With health information so easily available online, double-checking what a doctor says is much more common than before. In fact, 84% of patients verify the veracity of their physician's assertions after an appointment [Lau16]. The information they find may jeopardise their trust in their doctors, which may be misguided since the health information seeking capabilities of the patients are not always accurate (see Section 3.2 for more information on that).

Regardless, any account of a patient's trust on a physician has to take into account not only the interpersonal trust between them, but also the patient's social trust in collective health institutions, which may be influenced by the media and social bias [PR00].

3.1.2 Communication

Regardless of how a physician perceives online health information, they should develop their communication skills in order to debate it with the patient [MLP+03]. Moreover, they should listen to the patient's account without interrupting them, so that the patient can establish their identity and develop a stronger relationship with their physician. By not disturbing the narrative, the physician can also understand the patient's experience better, which will lead to a better treatment [Aro03]. But they should not shy away from asking questions either, since the patients will often omit information like their concerns and feelings that, unbeknownst to them, may be important to their healthcare [Aro03].

In a 1993 study involving 97 women with breast cancer, 49.5% of them had difficulty understanding their clinicians [Aro03]. Moreover, health professionals have been shown to underestimate the amount of information wanted by the patients and their understanding of information they receive, while overestimating the amount of information they share with the patients [HL10].

3.1.3 Improving the Relationship

The onus of improving the physician-patient relationship is on the former. To do it, they have to address the patient's concerns, ensure that they understand the information given, and be affectionate. The shared decision-making process hinges on the health professional showing openness to the patient's contributions and opinions, as well as sharing responsibility and control in the choice of treatments [Aro03].

Personalised social network systems and social dialogue systems have been shown to improve the physician-patient relationship and the latter's health knowledge. Discussion between both through technology is therefore an effective knowledge transfer tool and enables patients to better understand their medical conditions [HLS15].

3.2 Health Information Seeking

Health consumers want access to health information, being particularly fond of visual methods such as health videos with simple language that explain their health conditions [Lau16]. But even though around one in every twenty Google searches are healthcare-related [Lau16], not all consumers can successfully retrieve health information [FF03]. The low health literacy of the users and the use of different terminologies between them and their doctors (who may use more medical jargon) results in them not always finding the answers they want [Sou16]. Also, health information on the Internet may be inaccurate, misleading, or misinterpreted, which may result in patients making inappropriate clinical requests to their doctors [MLP+03].

Murray et al. [MLP+03] point out that health information on the Internet could be a great health equalizer if everyone had the same access to it. However, when the access, processing, and application of information is skewed towards the higher socio-economic groups, then it widens the gap between them and the more vulnerable individuals. The improvement of an individual's health information seeking capabilities is directly linked to their health literacy levels, as explained in Chapter 2. According to Murray et al. [MLP+03], improving the health information retrieving skills of the population should result from a combination of government and private efforts to teach information search and assessment skills. Those are also in line with the measures analysed in the aforementioned chapter.

The Internet seems to be an important factor in improving the health information and services patients receive [FF03]. However, they can also get the information they need from other people. In Section 3.1 we described the interactions between health professionals and patients but, when a health concern involves personal issues related to the endurance of a health problem or fast relief, patients tend to prefer non-professional contacts such as fellow patients, friends and family [Fox11]. This links again to Chapter 2, where the connections between communities, health literacy, and health information acquisition are explained in detail.

3.3 Personal Information Management

In the context of this work, the term "personal information" conveys the type of information an individual keeps for their personal use. According to Jones [Jon07], personal information management (PIM) may be defined as the acquisition, creation, storage, organisation, maintaining, retrieval, use, and distribution of information needed to complete certain tasks. However, he mentions that other definitions may include the person's methods for acquiring, organising, storing, and retrieving information. What all definitions have in common is the focus on how the information is arranged in order to be easier to find.

The main reason for the use of PIM methods is being able to check any information we might need during certain situations that may demand it. By using tools, technologies and techniques to help us spend less time with information management tasks, we gain more time to use that information smartly and innovatively for its intended purpose, as well as more free time for ourselves [Jon07].

The process of finding previously experienced personal information starts by recognising the necessity of looking for the information, followed by recalling details that allow the identification of the intended piece of information, recognising it, and repeating the process until all the necessary data is retrieved [Jon07]. Each of those steps is susceptible to failures, among which are information decentralisation (due to being spread across different locations, devices, forms, and associated institutions); the lack of foresight to know when and where to register information; the forgetfulness of previously stored information; and the bewilderment when confronted with information with an unfamiliar jargon or of extensive volume [PUC+06].

The concept of PIM is used in several fields, including healthcare, since consumers need ways of managing the health information they acquire. However, generic PIM tools may not support personal health information management (PHIM) requirements such as grouping information according to the corresponding phase of treatment (e.g. surgery or therapy) [PUC+06].

3.3.1 Applicability in Healthcare

Taking notes on paper is probably one of the oldest forms of managing information, but its flaws have long been recognised. Thankfully, the increasing interest in PIM methods and technologies has been prolific in creating new solutions [TJB06], even though some might increase the overall complexity of PIM by adding extraneous features that may difficult their use, as well as leading to information fragmentation [Jon07]. In healthcare, PHIM methods sometimes fail due to human error such as failing to record information correctly, or making mistakes and omissions when taking notes some time after an appointment [Gur12].

Regardless of where the technologies take us, it is the belief of Robinson et al. [RPE+98] that the health professionals' response to them are crucial to their adoption. They should make an effort to understand new methods, validate them and contribute to their use by the patients. Otherwise, the latter could be at a disadvantage without the knowledge and guidance the technologies could bring.

3.4 Summary

A patient can acquire information from different ways: their doctors, their community, and other sources such as the Internet. None of them are flawless: the doctors may not relay all the information they possess nor do it in layman's terms; the community's ability to help is directly linked to their health literacy levels; and the Internet is not always a trustworthy source of information.

Even when the patient successfully gains new health information, they have to manage it correctly in order to enable them to quickly retrieve it at a later date. That system of organisation is called personal health information management, and usually fails due to human errors, but its application in technology has not been perfect either. A crucial point for the technologies to be accepted is the physician's response to them.

Chapter 4

Technology for Health Consumers

4.1 Speech Recognition

Automatic speech recognition (ASR), often referred solely as speech recognition, refers to a process in which the voice of a human is heard through a microphone and subsequently examined in order to recognise every spoken word [RM13]. It encapsulates two different fields in itself: the acoustic model, which identifies language units such as phonemes present in the audio input; and the language model, which associates a probability to each word in the dictionary in order to better determine what is being said in each situation. Speech recognition systems have been used for several years and can have different uses, from dictation, to automated translations, or voice interfaces. When a speech recognition program is used to transcribe human speech, it can be called speech-to-text (STT).

There are several aspects that may limit the effectiveness of an ASR system. Gur [Gur12] points out that spoken language is more unintelligible than dictated language due to differences in the proximity to the microphone and succinctness; quieter backgrounds help the software better understand the speaker; some programs may learn from past experiences and hence be more adapted to specific speaking patterns and accents; and many ASR solutions do not differentiate between speakers, which is not ideal in a conversational scenario.

According to Reddy and Mahender [RM13], the process of recognising human speech is composed of 4 interconnected procedures:

1. Feature extraction, a method of creating evenly distributed discrete vectors of speech characteristics (called observations) from the sample audio;

2. Acoustic models database, a collection of recognisable vectors gathered from the training data which is compared against the extracted vectors in order to identify the sounds being made;

3. Dictionary, which makes the connection between the acoustic models identified in the speech with human vocabulary in the chosen language; and

4. Language model, a component responsible for checking the list of words that may match the speech in order to select the most likely to have been said according to context, grammatical rules, and statistical information.

The most noticeable difference between speech recognition systems lies in the speech recognition algorithm used. Modern ASR solutions use hidden Markov Models (HMM), which Rabiner and Juang [RJ86] define as "a doubly stochastic process with an underlying stochastic process that is *not* observable (it is hidden), but can only be observed through another set of stochastic processes that produce the sequence of observed symbols". They give the example of a hidden coin toss, in which the observer does not know how many coins are being tossed, or any other component of the experiment; they only have access to the result of the total coin tosses (head or tails) and, from that data alone, have to explain how those were obtained. The simplest solution would be a fair coin model, in which the coins used are balanced, there are two possible states for each, and each state correlates to a specific result (head or tails). However, several different models could have been used to obtain the same statistical results, ranging from different coin balances, to a different number of coins, or more complex states. The HMM are labelled as "hidden" exactly because the analysis is made without full knowledge of the data-gathering conditions, and therefore heavily rely on the amount of model training data given.

Hidden Markov Models are widely used in speech recognition because each speech signal can be interpreted as a stochastic process. Each word in the system's dictionary is modelled by different HMM, and the algorithm can then automatically train with a small data set [RJ93]. However, other techniques have been emerging, such as neural networks and, most recently, end-to-end ASR. Neural networks have been successfully paired with HMM methodologies in the tandem speech recognition system by applying discriminative feature processing before the HMM's Gaussian-mixture distribution modelling [HES00], and also by reducing the dimensions of complex data while preserving its characteristics [HZ10], diminishing the error rates of pure HMM models. End-to-end ASR does not combine neural networks methodologies with existing HMM systems like the previous examples; instead, it uses a recurrent neural network exclusively. Its biggest shortcoming is its reliance on an extra language model to correct the output, since it often presents spelling mistakes [GJ14].

4.1.1 History

Automatic speech recognition has been the culmination of centuries-long conceptualisation and technological advances. It can be argued that it started in the XIX century, when Alexander Graham Bell, Chichester Bell, and Charles Summer Tainter created a machine that would record human speech in a wax-covered cylinder with grooves, which would later evolve into the "dictaphone". This product was commercialised as a way of recording dictations from white-collar workers which could later be heard and transcribed by secretaries, reducing the costs of hiring stenographers [JR04]. Even though there was no actual ASR involved, this kind of dictation device illustrates how the automation needs felt more than a hundred years ago led to technological breakthroughs in that field.

The first ASR system was created in 1952 by Davis, Biddulph, and Balashek of Bell Laboratories; their solution recognised isolated digits according to the frequencies of the recorded voice. Other similar solutions such as vowel recognisers emerged in that decade; however, none of them were on-par with current ASR standards since they were all configured to receive only one digit or vowel at a time. The first system that used a speech segmenter to analyse the type of input was developed by Sakai and Doshita in 1962. Just a few years before, in 1959, Fry and Denes created the first system that used statistical data to determine allowable phoneme sequences according to the rules of the English language [JR04]. Both of these studies ended up shaping the design of modern ASR systems.

In the 60's and 70's, influential science fiction films such as "2001: A Space Odyssey" (1968) and the "Star Wars" original trilogy (1977-1983) revitalised the interest in ASR. They both showed automated entities capable of full ASR and speech modulation, highlighting the potential applications of such systems [JR04].

In the 1980's, the ASR vision changed to a more statistical modelling system instead of a template-based one, and the hidden Markov Model (explained in Section 4.1) gained prominence in the field. Neural networks were also introduced in ASR technologies around the same time [JR04].

More recently, call centres' interest in recognising the department to which a call should be redirected led them to create an ASR solution that would do just that. AT&T's Voice Recognition Call Processing, for example, was introduced in 1992 and currently handles around 1.2 billion calls per year. This and other similar solutions were only possible because great progress in ASR technologies was made during that decade, as systems became more sophisticated and able to handle more processing power and data [JR04].

Nowadays, speech recognition exists in several different ways around us, from simple transcription and translation applications, to intelligent home assistants such as Amazon Echo or Google Assistant.

4.1.2 Applicability in Healthcare

The use of a speech interface in healthcare may help eliminate spelling mistakes, alternative spellings, and abbreviations that could cause confusion due to the specificity and complexity of medical jargon [LTH+11]. It also helps people who type slower by automating that job and increasing their efficiency [JLL+14].

Several studies have been made in regards to the applicability of speech recognition technology in healthcare [Gur12; LTH+11; Rod13; FAR+13; Lob13; PKS04; SZH+15]. Their scopes are extremely varied, from apps that support the elderly, to real-time translations during medical interviews, clinical handover aids, or health advisors. Despite the diverse scopes, some conclusions can be drawn from those projects.

First of all, ASR makes less grammatical errors than humans when writing transcriptions, even though it is not as accurate overall; it is cost effective and significantly improves patient outcomes such as turnaround times [JLL+14].

Devine et al. [DGC00] compared, in 2000, three different products: IBM ViaVoice 98 (discontinued in 2005), Dragon Systems NaturallySpeaking, and L&H Voice Xpress. Their results show that the IBM solution had the lowest mean error rate for vocabulary recognition, at 7 to 9.1%, for general English vocabulary, medical abbreviations, and numbers. However, those trials were made by reading a script, so naturally occurring speech characteristics (such as hesitations and filler sounds) were not taken into account.

On the other hand, Liu et al. [LTH+11] did a similar study in 2011, focused on spoken clinical questions. To that effect, they used Nuance Dragon Gen and SRI Gen, both generic versions of ASR solutions; as well as Nuance Med, a medical version of the first solution. They also modified the SRI solution through a language model approach, calling the result SRI Adapted, which was developed specifically to better recognise medical speech. Their conclusions were that none of the original systems performed well with the clinical questions they were given. The best was SRI, with a 41.5% error rate. However, SRI Adapted had an error rate of 26.7%, significantly better than its unmodified version. A good example on how the two differed is the term “Paget’s disease”. While the second word is present in all dictionaries, the first one is usually misinterpreted due to its specificity to the medical field. Since SRI Adapted was given a domain-specific adaptation, it was configured to recognise the first word based on the successful interpretation of the second word.

It is important to notice that the considerable discrepancies between the results of both studies can be attributed to different definitions of error rates. For Liu et al. [LTH+11], the error rate is the “report [of] the sum of substitution, insertion, and deletion errors (referred to collectively as word error rates (WER), an established metric for evaluating speech-recognition systems)”. On the other hand, Devine et al. [DGC00] only considered substitutions and deletions when considering their error rates, which may justify their lower values.

4.2 Mobile Applications

Nowadays there are several technologies that try to negate the adverse effects of a low health literacy, bad physician-patient relationships, and inadequate PHIM. However, not all at-risk groups are equally represented in the app market, as can be seen in Figure 4.1.

Even though healthcare apps tend to be well received by physicians and patients alike, methods to evaluate them are lacking. Often doctors are hesitant to recommend a specific app and health consumers have to rely on the app’s rating, which may not be representative of the app’s quality but rather of other users’ experiences with it, which is a subjective metric [SDN+16]. Singh et al. [SDN+16] also noticed other problems with the apps: most are not secure, one third don’t inform their users on privacy issues, and many are often conceived as a paper alternative, which means that they don't react to irregular input such as levels of blood sugar above the recommended amount, for example, nor do they encourage the user to engage in healthy behaviours [Bro16].

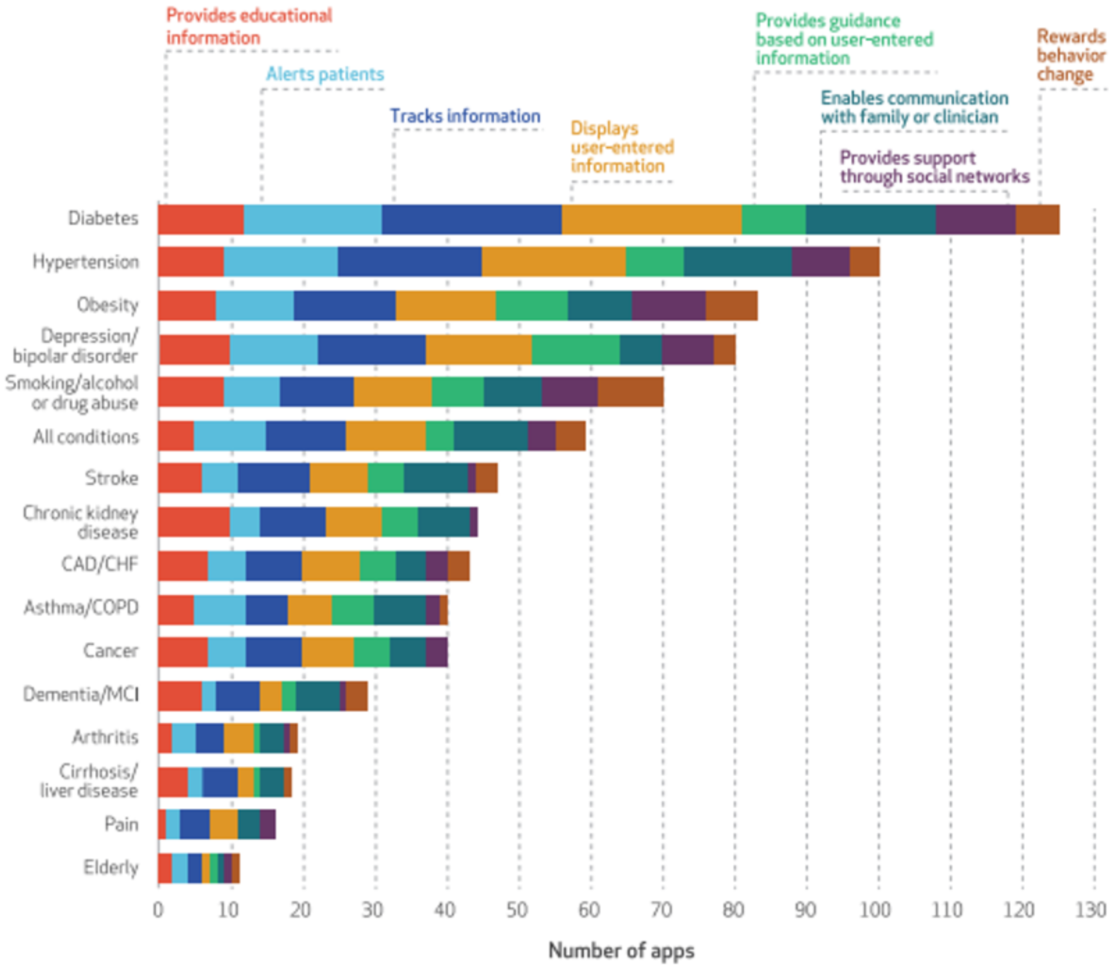


Figure 4.1: "Patient engagement–related functionalities of selected [mobile health-related] apps, by high-need, high-cost population, 2014–15." [SDN+16]

4.2.1 MyFitnessPal

MyFitnessPal [MyF17] is both a Web and mobile app for iOS and Android. It is free to use, but includes in-app purchases of 0.79 Euro up to 56.20 Euro. It is very widespread, with more than 50 million downloads. Its popularity might have to do with its compatibility with external devices such as smartbands and complementary apps, which make this app a more versatile choice and more inclusive of different user groups.

Its functionalities are mostly based on calorie counting, both the ingested calories through detailed logs of food consumption, and the spent calories by means of an exercise calculator and instigator. A study by Luna et al. [LGF+15] discovered that a group of people questioned before and after using this app for 15 days actually learned information that is shown passively in the menus. The subjects recognised with much more ease the families to which certain food products belonged to after using the app, and that information is shown in the screen associated with each food product when the user is listing their calorie consumption habits. Therefore, it was information that was not actively looked for by the participants, but ended up being learnt by them.

On top of its aforementioned functionalities, it also works as a PHIM system by saving calories and weight over large periods of time and showing that information back to the user whenever it is requested. This app is available in 21 different languages, including English and Brazilian Portuguese.

4.2.2 Patient Journey App

Patient Journey App [Hea17] is an Android and iOS mobile app developed by Health Care Labs B.V. that works both as a communication tool between physicians and patients and as an educational information source.

It gives health professionals the tools to curate what information to show each patient and when to do it. That information is meant to be both instructional and educational, guiding the patient in their current step of treatment. The physician can also set reminders in the form of push notifications for specific issues like taking a certain medication or carrying out certain actions (such as fasting before a surgery).

The app also supports direct communication between both parties, encouraging the patient to ask questions when having doubts and the physician to check-up regularly on them. The patient's family and friends can also be granted permission to access the patient's information timeline.

Throughout the whole process, the physician has access to a dashboard that shows statistics on how engaged the patients are, so that the physician can adjust the flow of information accordingly.

The app is free of charge for patients, but has to be first acquired by the health institutions in order to personalise it and grant them access to the back-office functionalities. The price they have to pay is unknown. The app is available in Dutch, English, French, Italian, Japanese, German, and Spanish.

4.2.3 Patient.info

Patient.info, or Patient.co.uk [PE16], is both a Web and mobile app (supported in Android and iOS devices) with an extensive database of medical conditions, medical services (from general practitioners to dentists, pharmacies, hospitals, and opticians), drug information, medical calculators, decision aids, clinical guidelines, health checks, and, contingent upon the purchase of an upgrade, clinical articles and medicine leaflets. It was first launched in 1997 by Patient information Publications (PIP), but since 2002 it has also been supported by Egton Medical Information Systems Ltd (EMIS).

Its main goal is, just like their slogan advertises, to give "trusted medical information and support", aiming at giving patients more access to free and complete health information. The users can check their conditions by searching for any term or by grouping the results by symptoms, alphabetical order, or category. It is therefore focused on people with health conditions who want to understand more about the illness that afflicts them and the closest places they can go for help. Furthermore, the medical calculators and online checks can also instil and promote healthy behaviours in the users.

As mentioned before, the access to the available information is mostly free, with the aforementioned extras costing around 12 Euro each. The location services work exclusively in England, and the information can only be displayed in English.

4.2.4 Patient Access

Also developed by PIP and EMIS, Patient Access [PE15], formerly known as EMIS Access, is a Web platform and mobile app for Android and iOS devices that supports the scheduling of general practitioner appointments, the consultation of one's medical record, and the ordering of repeat prescriptions for patients from the United Kingdom. The patients find in the app an easier and faster method to do those tasks while keeping a personal health record that gathers data measured from Apple Health and can be shared with the patient's physician. This app is also available exclusively in English.

4.2.5 Other Apps

There are several more apps available that prompt users to lead better and healthier lifestyles by exercising, eating adequately, sleeping better, and being generally more self-caring. Among those we can highlight MySNS[1], Apple Health[2], S Health[3], C25K[4], Daily Yoga[5], ShopWell[6], Meditation Studio[7], HealthTap[8], and several others. They vary wildly in terms of information available and functionalities, being that many of them are not only health information apps, but also personal health information management systems and are sometimes associated with health assessment hardware such as smartwatches.

Those apps were not described in greater detail either because their scope falls outside the intended breadth of this dissertation or because they are alternatives to the apps already detailed in Sections 4.2.1 through 4.2.4. However, there is another type of apps that, even though not meant to be used exclusively for healthcare, is worth mentioning: note-taking apps. From those, we would like to highlight Evernote due to its pervasiveness in that market, as well as its use of STT technology.

[1] https://www.sns.gov.pt/apps/mysns/
[2] http://www.apple.com/ios/health/
[3] http://shealth.samsung.com/
[4] https://play.google.com/store/apps/details?id=com.c25k
[5] http://www.dailyyoga.com/
[6] http://www.shopwell.com/
[7] http://www.meditationstudioapp.com/
[8] https://www.healthtap.com/

4.2.5.1 Evernote

Evernote [Cor17] is a note-taking app with several added functionalities. It is supported online through Web, desktop, iOS, and Android apps. Its most fundamental functionality is taking written notes on a subject by specifying the post's title, contents, and tags, and by enabling simple formatting of the text and organisation through a folder system.

However, the app is also capable of saving web pages, storing pictures, parsing hand-written text, taking speech notes, and scanning documents. It is hence a multi-purpose tool that works as a PIM system and can be used in the healthcare area. However, its lack of focus on any specific field makes it more generic and less able to cope with specific information problems that may be prevalent in a particular area. Instead of expanding its services horizontally it chose to grow vertically, increasing its utilities across a plethora of related and useful extra functionalities.

Evernote's usage was tested in a scientific context by Walsh and Cho in 2013 [WC13] and was shown to be more efficient than traditional paper-based PIM methods. The biggest advantage of paper was the ability to freehand directly into a notebook, but, in the experiment, Evernote overcame that inadequacy with its extra functionalities that are not achievable without a digital medium.

It is free to use, but has an associated premium service which goes from a 29.99 Euro to a 59.99 Euro one-year subscription and adds some of the mentioned extra functionalities, customer support, extra storage, and more sharing options. It is available in 24 different languages, including English from the USA and both European and Brazilian Portuguese.

4.3 Summary

Automatic speech recognition has improved significantly over the last few decades, to the point where it is now commercially available to the public in several different solutions ranging a lot of functionalities. However, its efficiency at recognising human speech does not always apply to the medical field due to the specificity of the vocabulary used in that area. In this field, tests suggest that the inclusion of a dictionary of specific terms heavily influences their perception.

Several healthcare-related apps already exist. Even though some are more popular and used than others, they all have benefits and drawbacks that should be taken into consideration when creating a potential competitor. That was what we did in Section 7.1, where a comparison between HealthTalks and some of the aforementioned apps can be seen.

Chapter 5

Characterisation of the Target Population

Even though the previous chapters improve our understanding of situations like the doctor-patient relationship and a patient's PHIM habits, some of these problems have not been deeply studied in the Portuguese population. For that reason, we decided to survey Portuguese patients, mostly around the Porto region, in order to gain a better perspective on their specific reality and needs. We decided to focus on three major issues: 1) note-taking behaviours regarding personal health information; 2) relationship between patients and doctors and both parties' behaviours during a medical appointment; and 3) their interest in the functionalities we had planned for HealthTalks (more information on those can be found in Chapter 6).

All of the questions included in the questionnaire and the answer options are listed in Appendix B in their original Portuguese version.

5.1 Development

The questionnaire went through various stages over its development, summarised in Figure 5.1. We were especially concerned with the complexity of the language used, since we wanted to ensure its readability by people from different educational backgrounds. Its structure was revised several times in order to limit its complexity and increase its accessibility. Our intention was to obtain the most extensive and most varied sample possible, thereby representing more accurately the general population's characteristics. During that process we were helped by Dr. Dagmara Paiva, who suggested the rephrasing of some questions and answers.

After the first prototype was developed (a copy can be seen in Appendix A), we tested it with friends and family in order to get a better perception on its clarity and simplicity. We tested it with 8 different adults, with ages ranging from 29 to 91 years old, and with education levels ranging from less than 4$^{\text{th}}$ grade to a bachelor's degree. A complete characterisation of those testers can

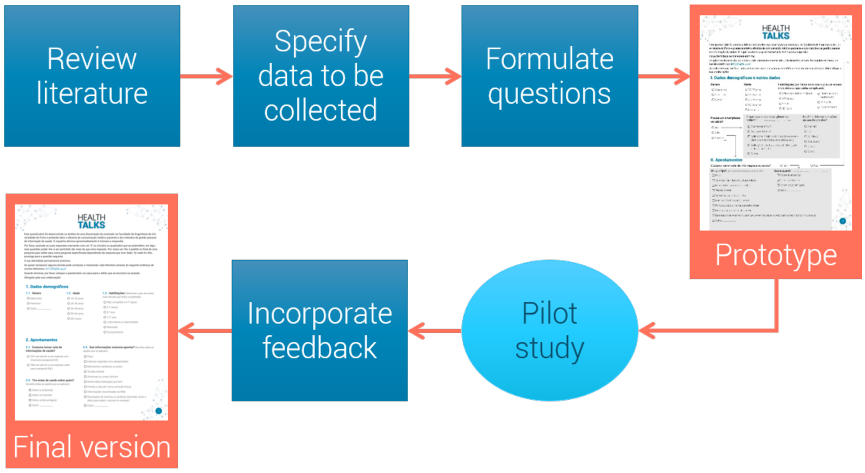


Figure 5.1: Process for creating the questionnaire.

be seen in Table 5.1. In addition, we also presented the questionnaire to two children aged 8 and 11, since their feedback was relevant in terms of their education level, even if they did not grasp some of the concepts due to their young age.

Table 5.1: Characterisation of the adult testers who participated in the pilot study.

		Qualifications				**TOTAL**
		<4	**4**	**12**	**B**	
Age bracket	**26-35**	-	1M	1F	-	2
	36-49	-	-	-	1F	1
	50-64	-	-	1F	1M	2
	65+	2F	1F	-	-	3
TOTAL		2	2	2	2	**8**

NOTE: The columns “<4”, “4”, “12”, and “B” stand respectively for “lower than 4^{th} grade”, “4^{th} grade”, “12^{th} grade”, and “bachelor’s degree”. “F” and “M” stand for “female” and “male”; the numbers before those letters represent the amount of people of that gender that are included in that specific category.

Our conclusions with that pilot study were very enlightening and led us to iterate on the prototype. We had used arrows to indicate dependencies between questions and jumps that the respondents had to make, but after those tests we understood that they were not very clear for many people, since most testers ignored them or were confused by their implementation. We replaced them with text indicating the question the respondent should answer next, according to the answer given, and wrote instructions in the beginning of the document to explain its non-linear structure.

Even though those were the most visible changes, other details were polished in accordance to the received feedback. Among those changes, we can highlight the unification of the questions’ layout structure to be consistent throughout the document (with the answer options always below the question, and never in-line with it); the increase in space between questions for legibility; the reorganisation of the questions to create a more cohesive flow of questioning and to capitalise on its non-linear structure; and the rewording of specific terms that were unknown to lower-educated testers.

All of the questions, excluding the last one, are close-ended in order to reduce the amount of effort necessary to answer them. Many of the multiple choice questions are ordinal scales with four options from "never" to "always"; we decided against an odd number of answer options to prompt the respondent to choose between a lower or higher frequency, avoiding neutral responses. For the multiple response answers we tried to create a comprehensive collection of answer options, but we always included the option "other" so that the respondent could relay their personal experiences. On the question about HealthTalks' functionalities and their importance, we used a Likert scale with just 3 items instead of the typical 5 or 7 [AS07], just from "disagree" to "agree", since we were interested in the overall opinion on the app and not in specific degrees of agreement.

Two identical variants of the questionnaire's final version were developed: one for printing and another for online distribution[1].

5.2 Online survey

The online variant of the survey was divulged through networks of friends and colleagues. The non-linear structure of the written copy was reproduced with Google Forms'[2] options to only display some of the questions under specific conditions, thereby streamlining the respondent's experience while answering and automatically avoiding invalid answer combinations.

This version was available to answer between January 11th and February 18th 2017. Two copies of that same survey were publicised in different ways. The first was shared on social media platforms and one symposium (DSIE, on January 31st), while the second was sent by email to most students and some staff of the University of Porto (UP). The total reach of the first post is hard to measure, since it was shared by third parties in private and public groups, but we know that the second version was received by 34727 people, of which 348, or approximately 1%, were staff, and the other 99% students.

The first survey was answered by 127 people and the second by 943, for a total of 1070 answers. We chose to develop different versions of the same survey to have the flexibility of differentiating the source of the answers during the data analysis. However, due to the proximity of both sets' data and the considerate difference in sample size, we decided to merge them together for the purpose of this analysis, treating all 1070 answers equally.

The raw data was processed in order to follow certain specifications, and was then interpreted in R. The only question that was not handled in this manner was the last one, which was open-ended and inquired what other functionalities the respondent would add to HealthTalks, or how they wished the app worked. Those answers were read individually and manually tagged with one or more topics. Those topics were initially very granular, but were later grouped in a broader

[1] Available at https://goo.gl/forms/2cEtihKnloWaCxdD3 (Portuguese only)
[2] https://docs.google.com/forms/

classification system in order to understand the general tendencies and desires of the respondents. For example, at the beginning we made the distinction between “I want to be able to send messages to my physician”, “I want to be able to contact a different physician in order to get a second opinion” and “I want real-time communication with any physician”. In the end, however, all of those use cases were combined into the topic “Communication with physician” since that is the idea that all of them share, even if in slightly different contexts. We added an explanation to each topic in order to convey the range of requests received for each of them.

5.2.1 Sample characterisation

In Figure 5.2 there are distributions of the respondents according to gender, age, and qualifications. The typical respondent (mode of the results) is an 18 to 25 year-old woman who completed the 12^{th} grade (31% of the answers correspond to this profile).

The discrepancies seen in the figure can be understood when taking into account that at least 88% of the total answers have been given by people who are in a higher education environment (either students or staff), and the other questionnaire was also originally mostly spread among people aged 18 to 25. The only significant discrepancy for which there is no clear reason is the one between genders, since the distribution of women and men in the Portuguese society is 52% to 48% [Ins12].

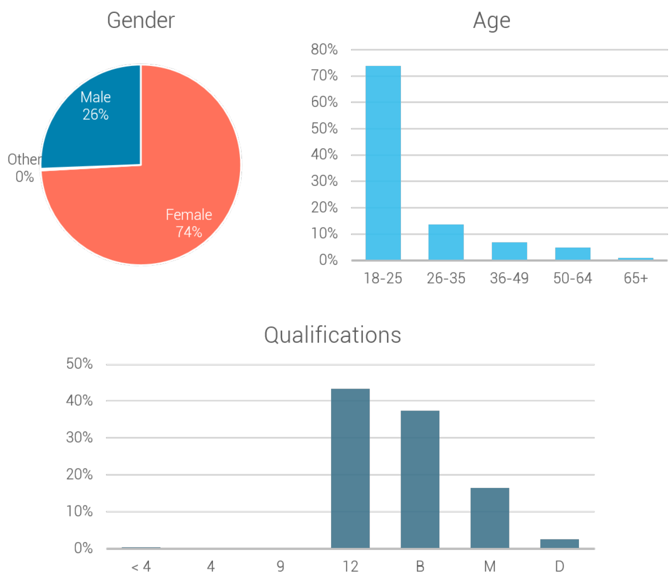


Figure 5.2: Distributions of gender, age, and qualifications in the online survey. In the qualifications graph, each bar corresponds respectively to: “lower than 4^{th} grade”, “4^{th} grade”, “9^{th} grade”, “12^{th} grade”, “bachelor’s degree”, “master’s degree”, and “doctoral degree”.

5.2.2 Generic Data Analysis

We decided to combine the data from the questionnaires into graphs in order to make it more understandable at a glance. In Figure 5.3 we can see that 61% of respondents do not take health notes; those who do so tend to write about themselves, on a piece of paper, to register their weight. Those are the most common answers to these questions, though people appear to take health notes about several topics.

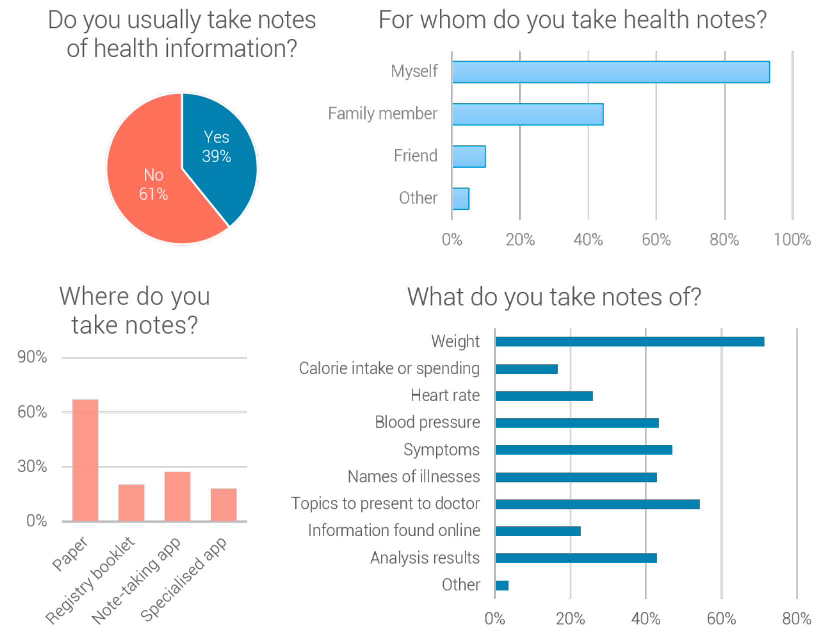


Figure 5.3: Note-taking habits of health information of the online survey's respondents.

In Figure 5.4 we can see that most people never take notes during an appointment, or do so on occasion. Almost half of the people who take notes of a consultation do so right after it finishes. In terms of information written, the treatment and dosage are the most common, followed by the prescribed medication.

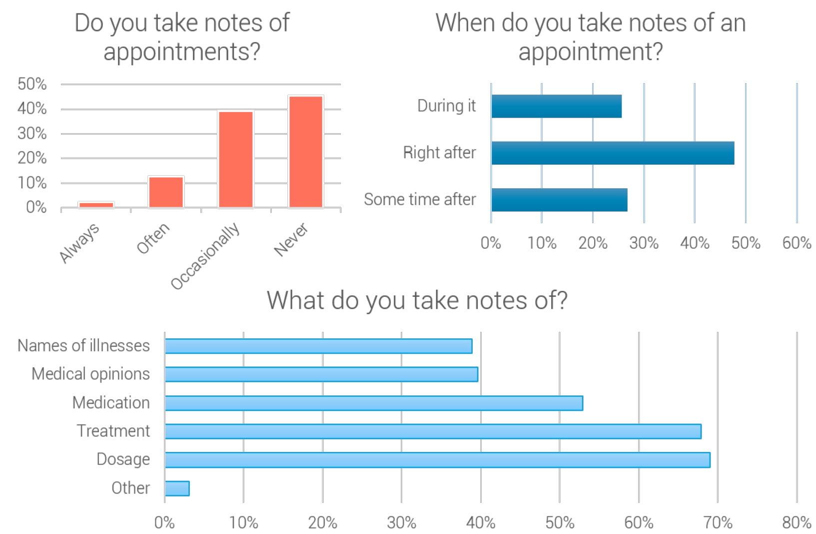


Figure 5.4: Note-taking habits of the online survey's respondents during an appointment.

The graphs in Figure 5.5 are very similar, and both answer "occasionally" to the questions on how often they forget something the doctor said; and if their physician gives them leaflets, notes, or other informative material in paper.

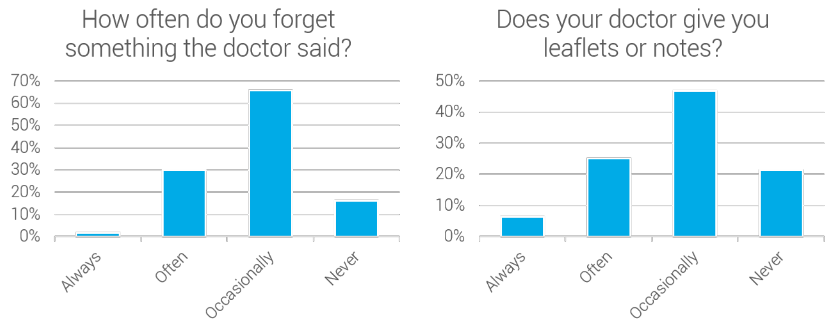


Figure 5.5: Graphs detailing how often a patient forgets what their doctor said, and how often the latter give notes to their patients, according to the online survey.

When looking at the graphs in Figure 5.6 we can understand a clear difficulty with patients in understanding their doctors due to their use of medical terms. However, most respondents always or often understand their doctors. When they do not understand what the doctor said, the vast majority always or often asks them to repeat, but those who do not appear to be moved by not wanting to be a bother or seem ignorant.

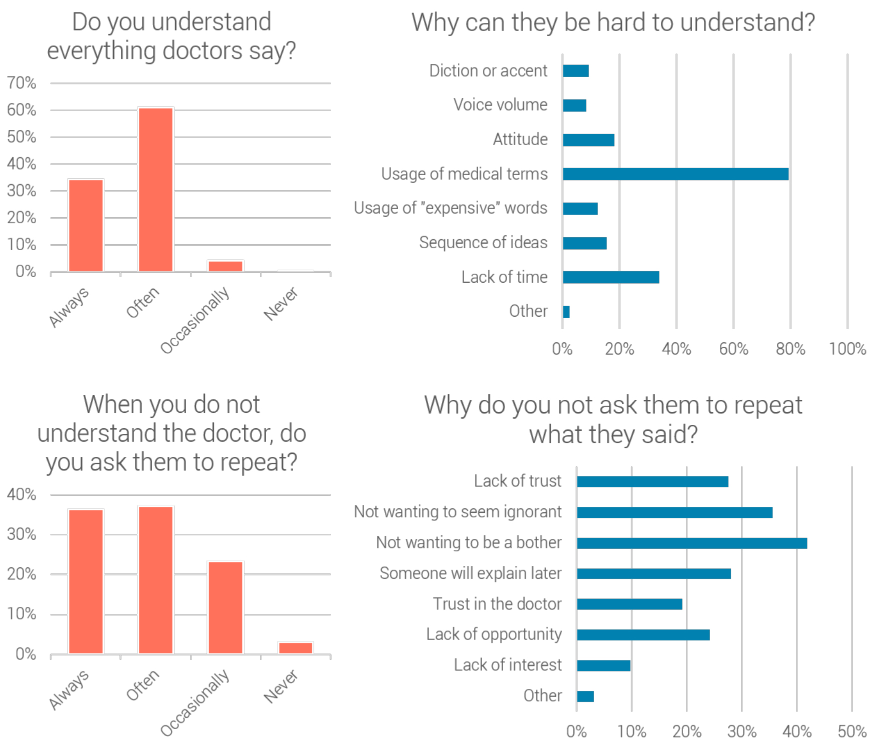


Figure 5.6: Graphs regarding the communication between patient and doctor, as answered in the online survey.

Looking to the answers about mobile devices (see Figure 5.7), we can see that almost 95% of respondents have a smartphone or tablet. In terms of usage, it appears that most actions listed (calls and texts, internet access, and apps) are used thoroughly. When it comes to operating system, almost 70% are Android, followed by iOS with almost 30%.

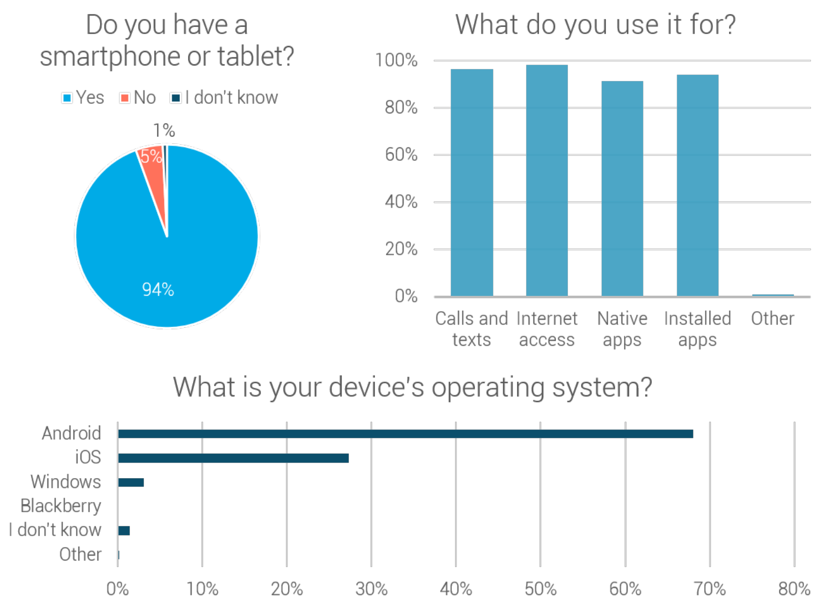


Figure 5.7: Graphs regarding the ownership and usage of mobile devices, as answered in the online survey.

HealthTalks' speech-to-text functionality appears to be the one considered the least important, according to the results in Figure 5.8. In contrast, both the definitions of medical terms and the possibility of taking notes are considered important by the vast majority of respondents. Overall, 38% of respondents would try HealthTalks, with 46% more undecided.

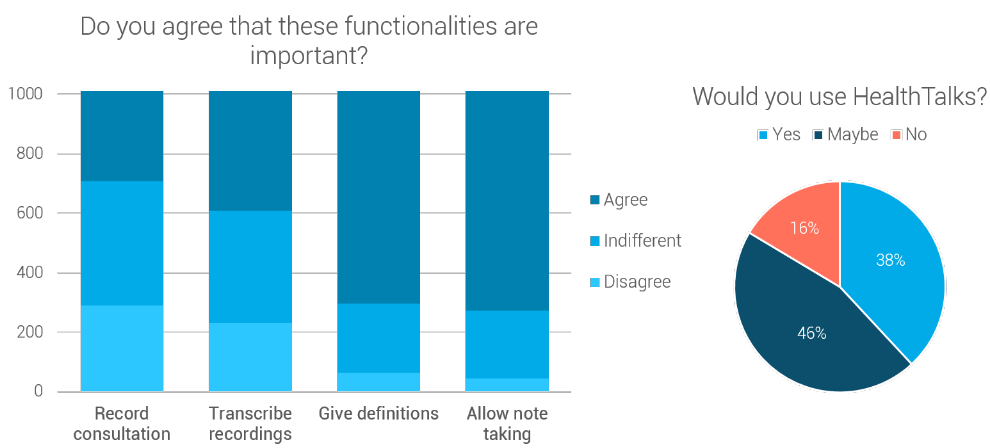


Figure 5.8: Graphs regarding the appraisal of HealthTalks' features, as answered in the online survey.

5.2.3 Gender Analysis

All of the answers were analysed taking into account the demographic characteristics of the respondents in order to detect any relevant correlations that could change our perception of HealthTalks' target population. To analyse the association of gender with each variable we used the Chi-squared test; in Table 5.2 are the significant differences ($p<0.05$) between genders. We did so by comparing the proportion of positive answers to the questions between the genders. The option "other" regarding gender was not considered since it had too few occurrences to extrapolate from.

Table 5.2: Association between gender and other answers in the online survey.

Answer	Male	Female	χ^2	p-value
I take health notes	0.29	0.43	17.45	2.96e−5
I take health notes and register weight	0.83	0.68	6.87	8.79e−3
I take health notes and register heart rate	0.40	0.23	9.58	1.97e−3
I take health notes and register blood pressure	0.55	0.41	5.37	2.05e−2
I take health notes and register points to discuss with the doctor in a future appointment	0.40	0.58	8.07	4.49e−3
I take notes of an appointment and register medical opinions	0.49	0.37	6.08	1.37e−2
I do not ask the doctor to repeat something I did not understand because I trust them	0.34	0.15	20.33	6.52e−6
I do not ask the doctor to repeat something I did not understand because I am uninterested	0.19	0.07	13.11	2.94e−4

It can be seen that women take more health notes than men and focus more on writing points to discuss in a future appointment, whereas men, even if taking less health notes, are more likely to register weight, heart rate and blood pressure, though it is unclear if it is their own or a family member's or friend's. Men are also significantly more likely to not ask the doctor to repeat something due to trusting them or being uninterested in the message. The trustworthiness of doctors is also seen in the answer about registering medical opinions from their physician.

5.2.4 Age Analysis

Just as in the previous section, we also associated the answers given by the respondents with their age bracket to see what correlations we could find. The significant results of the Chi-squared test and respective p-values can be seen below in Table 5.3. Please note that the option “65+ years old” was not considered due to appearing too infrequently.

Table 5.3: Association between age and other answers in the online survey.

Answer	18-25	26-35	36-49	50-64	χ^2	p-value
I take health notes	0.35	0.48	0.53	0.48	18.27	3.88e−4
I take health notes and register the information of a family member	0.40	0.43	0.64	0.64	12.71	5.31e−3
I take health notes and register weight	0.76	0.66	0.62	0.52	10.32	1.60e−2
I take notes of an appointment and register the prescribed medication	0.59	0.42	0.40	0.25	23.90	2.62e−5
I take notes of an appointment and register the dosage	0.75	0.58	0.50	0.53	23.46	3.25e−5

By analysing the results, we can ascertain some trends such as the increased likelihood of a person registering health information of a family member as they get older (which is justifiable by possibly having an older partner, older parents, and children or grandchildren to care for). Another clear trend is that younger people are more likely to register their weight, which may potentially be more due to body image issues than health problems. Also more common among younger people are the habits of registering the prescribed medication and its dosage, which is a counter-intuitive result since older people tend to take more drugs [RL]. One possible justification is that older patients might depend on other means to remember that information (such as the prescription given by the physician). In general, people aged 36 to 49 years old are the most likely to take health notes, which may be related to having young children (in 2013, the mean age of mothers giving birth in the USA was 26 years and presented an upward trend [MHO+15]).

5.2.5 Qualifications Analysis

Lastly, we tried to find parallels between the qualifications of the patients and their answers. The results of the chi-squared test in which the p-value was lower than 0.05 can be found in Table 5.4 below. We disregarded the options “less than 4th grade”, “4th grade” and “9th grade” since too few respondents selected them.

Qualifications are inevitably linked to age, since completing higher degrees may require several years and they have to be done in sequence. However, even though more qualified individuals are less likely to be very young, less qualified people can be of any age. Hence, it is surprising that for responses such as taking health notes, taking health notes of a family member, and the registering of the prescribed medication, the trend seen in regards to qualifications is very similar

Table 5.4: Association between qualifications and other answers in the online survey.

Answer	12	B	M	D	χ^2	p-value
I take health notes	0.34	0.44	0.52	0.48	10.50	1.47e−2
I take health notes and register the information of a family member	0.36	0.45	0.55	0.77	13.78	3.21e−3
I take health notes and register blood pressure	0.35	0.45	0.55	0.46	8.88	3.10e−2
I take notes of an appointment and register the prescribed medication	0.65	0.47	0.41	0.41	24.01	2.48e−5

NOTE: The columns “12”, “B”, “M”, and “D” stand respectively for “12$^{\text{th}}$ grade”, “bachelor’s degree”, “master’s degree”, and “doctoral degree”.

to the one seen in regards to age, but in this case with more qualified people registering more information of family members and being less likely to register the prescribed medication, while the habit of taking health notes is more prevalent in people with a master’s degree. Our interpretation is that since the population queried was generally highly educated and affiliated with a university, their academic achievements did relate directly to their age.

The only unique variable is the registering of blood pressure, which is slightly more prevalent in people with a master’s degree.

5.2.6 Suggestions Analysis

When asked what other functionalities they would add to HealthTalks, or how they wished it would work, the answers were surprisingly consistent among respondents, with two topics consistently being mentioned: the addition of reminders and a calendar. 237 people answered the question (more than 1 answer for each 5 respondents), though not all were valid responses. As shown in Table 5.5, the final grouping of answers resulted in 12 non-exclusive topics (see Section 5.2 for more information on the creation of those topics).

Table 5.5: Topics mentioned in open-ended answers in the online questionnaire.

Topic	Explanation	Prevalence
Reminders	Addition of a reminder or alarm system to prompt the user to take medication, undergo treatments, or go to appointments and exams.	25%
Calendar	Implementation of a calendar to check appointment and exam dates, as well as periods of medication and treatment.	13%

Other functionalities	Suggestions that did not fit any topic. See below for examples.	12%
Existing functionalities	Suggestions or reinforcements of an idea that was already presented as being core to HealthTalks (recording, transcription, PHIM, or medical terms definitions).	12%
Health summary	A place where, at a glance, a person could understand their latest health developments, including appointments, medications, symptoms, and exam results.	11%
Communication with physician	A way of communicating with their doctor or a different one to be diagnosed or receive recommendations.	9%
Glossary	A persistent and searchable glossary of medical terms, medications, diseases, and any other relevant health information.	8%
Health tips	Suggestions and recommendations appropriate to several different situations (such as how to treat certain symptoms, how drugs interact, or messages that promote well-being).	8%
Health monitoring	Functionality to track the evolution of certain aspects over time (weight, exam results, medication intake, menstrual cycles), possibly in a graphic.	8%
Integration with existing systems	Having the possibility of retrieving a patient's medical history, or being integrated with the National Health Service, and scheduling appointments.	7%
Simplicity	The app has to be easy to use, especially by the elderly and uneducated patients.	4%
Ethics	Mention of the ethical problems posed by recording an appointment, or the potential privacy problems when storing that and other health information.	4%

In addition to those topics, 10% of all answers were invalid (by either not answering, giving a null response, or not mentioning any functionality or aspect of the app). The "other functionalities" mentioned by respondents included the association of exam results, prescriptions, photographs, or other media with a appointment; a forum where patients with the same illnesses could meet and exchange experiences; a map with the location of nearby pharmacies; video lectures on specific procedures and diseases; a built-in tool to translate the transcriptions to a different language; computer-generated summaries of medical appointments; and comparisons between a patient's data and that of the general population.

5.3 Printed survey

On January 3rd 2017 a request was made with the Health Ethics Committee (CES) of the São João Hospital Centre (CHSJ) in order to get permission to hand the questionnaires to patients inside that health establishment. The permission was granted on March 17th and we started distributing them on April 28th. Although we were initially told by CHSJ to deliver 500 copies a week, and we planned for a month of distribution by printing 2000 copies, a few weeks later we were told that, in similar situations, the CHSJ only recovers around 380 questionnaires for every 4000 they distribute. That is mostly due to the patients delivering the questionnaires in the wrong departments and the difficulty of the staff in quickly coordinating to correct those mistakes. Other questionnaires may be lost or thrown away by the patients.

The most recent collection of questionnaires was on June 23rd, adding to a total of 239 questionnaires. That is around 12% of the printed copies, but we will continue to recover questionnaires as they keep being delivered by the patients, in order to obtain an even more significant sample.

Since the questionnaires were delivered in print, the first phase of data processing consisted in introducing those results in a digital file to be easily analysed. That was done by copying the Google Forms questionnaire used for the online version and submitting these answers in that form. This allowed us to obtain a final file with the same formatting as before, so the rest of the methodology was similar to the one described in Section 5.2.

However, the online form had restrictions in place to control the flow of questions according to previous answers. Those restrictions were communicated to the respondents in the paper version, but many did not follow the instructions. Hence, 79 questionnaires (around a third of the total) were not analysed since they did not comply with the rules, resulting in 160 valid answers. One example of such non-compliance would be answering how useful one of HealthTalks' features is, but not the others; or not answering how often they do not understand their doctor, but then specifying a reason why they do not understand them. We chose not to analyse those 79 answers in order to ensure coherence between the analysis of both versions of the questionnaire.

5.3.1 Sample characterisation

The distributions of the respondents can be seen in Figure 5.9. Contrasting with the homogeneity of the online survey, the printed survey respondents are more diverse in terms of age and qualifications, even though 66% of them are women, which again is higher than the population average.

The amount of respondent diversity supports our initial belief that this version of the questionnaire would be more representative of the population at large.

5.3.2 Generic Data Analysis

Just as before with the online questionnaires, in this section we will analyse the answers to each question. The results to the first few questions can be seen in Figure 5.10. Most people take health notes, and when they do they are often about themselves or a family member and written

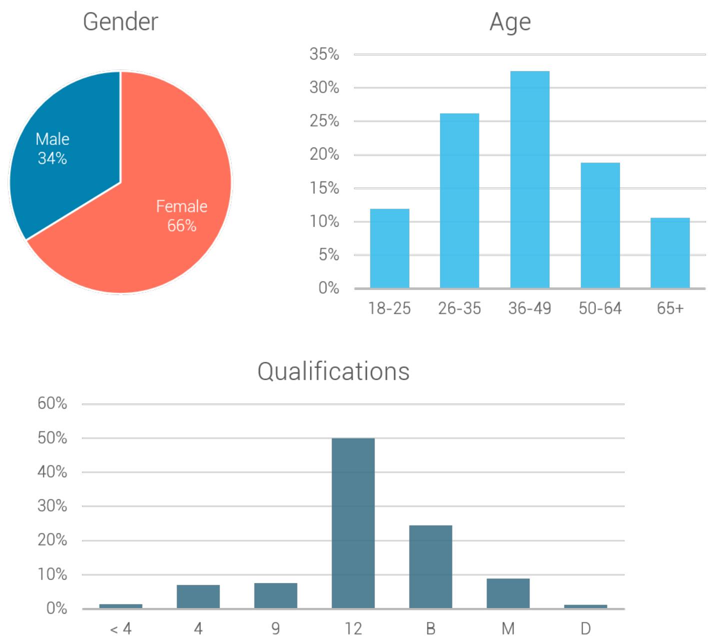


Figure 5.9: Distributions of gender, age, and qualifications in the printed survey. In the qualifications graph, each bar corresponds respectively to: “lower than 4th grade”, “4th grade”, “9th grade”, “12th grade”, “bachelor’s degree”, “master’s degree”, and “doctoral degree”.

on paper. In regards to health issues written about, the most common are weight, blood pressure, and analysis results; whereas the least common are calorie intake or spending, information found online, and heart rate.

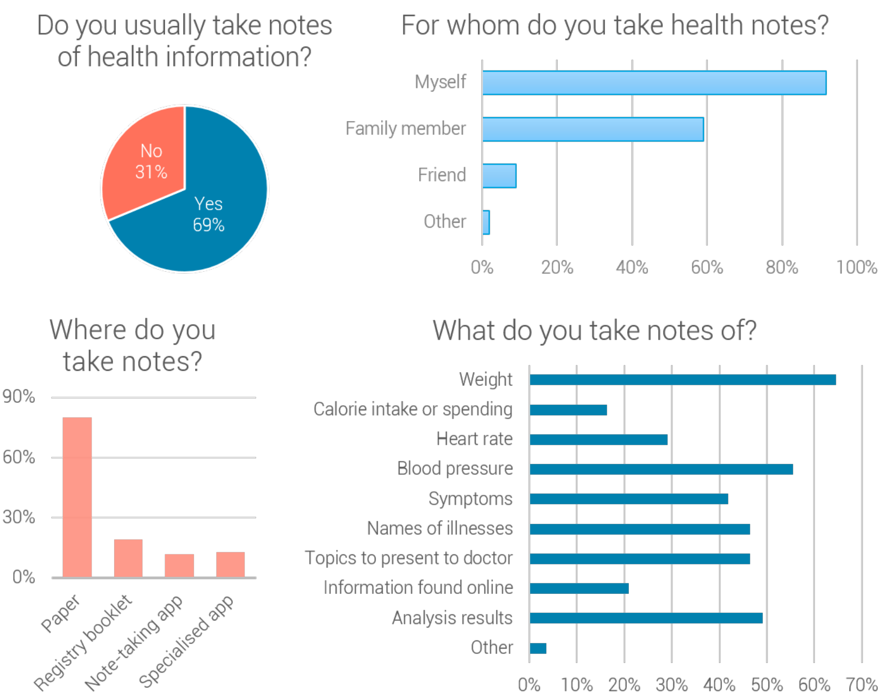


Figure 5.10: Note-taking habits of health information of the printed survey’s respondents.

When it comes to taking notes during a medical appointment, around 70% of people never or seldom do. Among those who take notes, the vast majority do not do it during the appointment, but rather immediately or some time afterwards. In terms of content written, it seems to be somewhat homogeneous, with treatment, medication, and dosage being slightly more mentioned. All of those statistics can be seen in Figure 5.11.

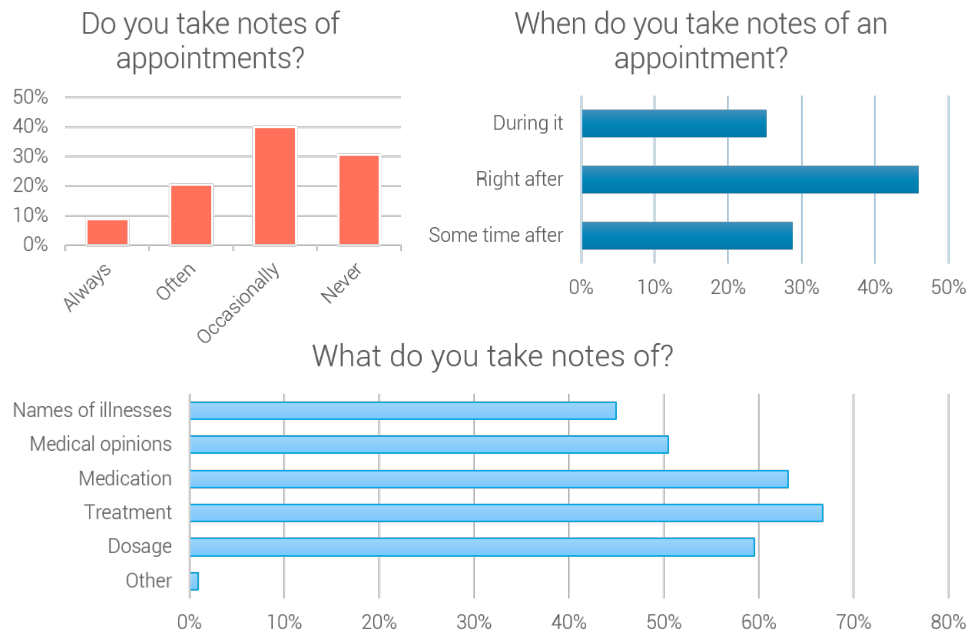


Figure 5.11: Note-taking habits of the printed survey's respondents during an appointment.

In Figure 5.12, we can see that most people occasionally forget what their physician says, and it is also only occasionally that the latter give some kind of paper support (be in a leaflet, note, summary, or similar) to their patients.

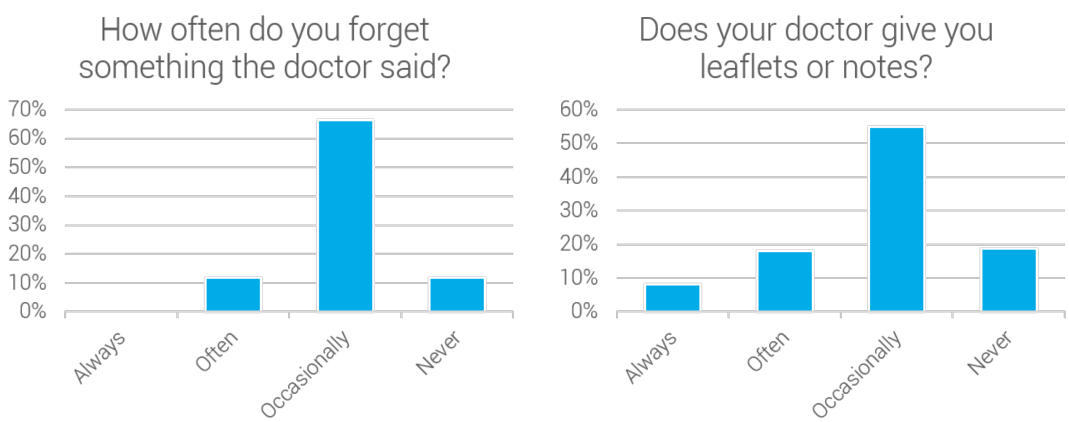


Figure 5.12: Graphs detailing how often a patient forgets what their doctor said, and how often the latter give notes to their patients, according to the printed survey.

Regarding the doctor-patient communication, it is clear that most patients do understand what their doctor says at least often, and with the same frequency ask them to repeat in case they missed it. However, in terms of the reasons why the doctor may be misunderstood, the use of medical terms is clearly the most common reason, as seen in Figure 5.13. In contrast, when asked why they do not ask their doctors to repeat themselves, patients are not as unanimous, though not wanting to be a bother or looking ignorant were chosen most often.

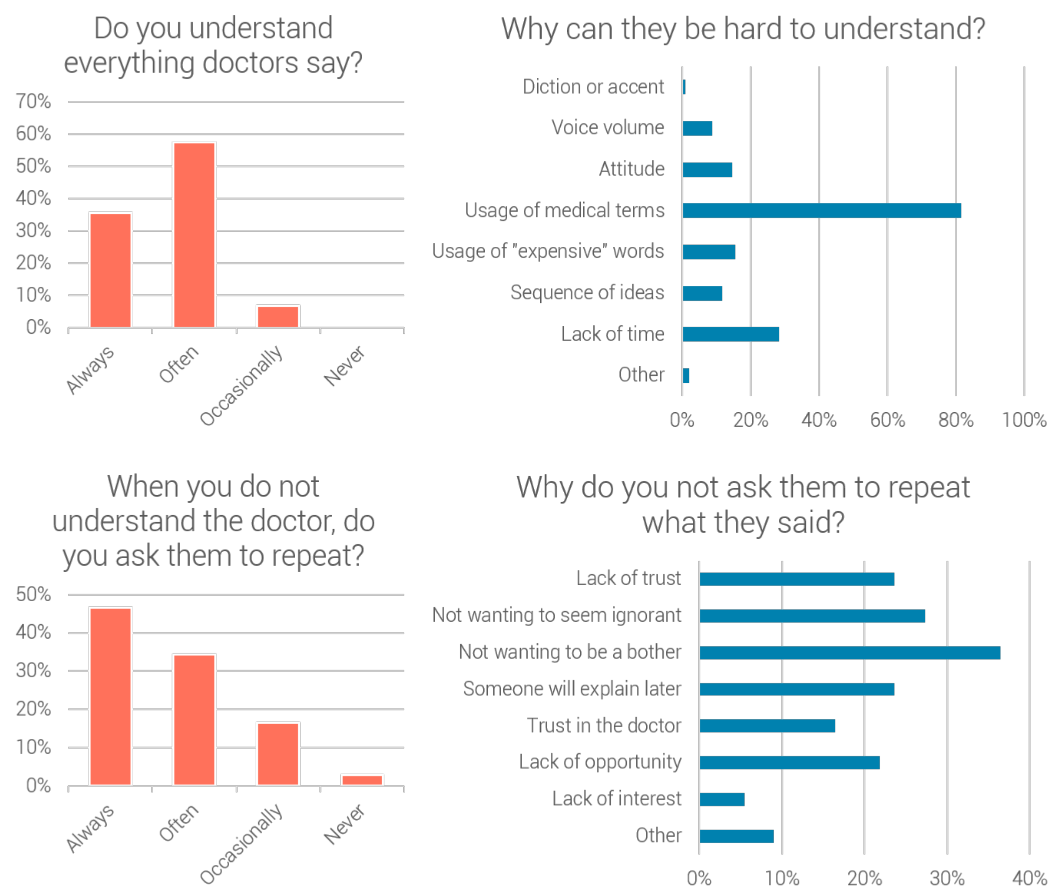


Figure 5.13: Graphs regarding the communication between patient and doctor, as answered in the printed survey.

When it comes to mobile devices, 84% of patients have one according to the graphs in Figure 5.14. Among those, more than 70% are Android devices. The usage of the device's functionalities is very uniform between the different categories, although internet access was the most common.

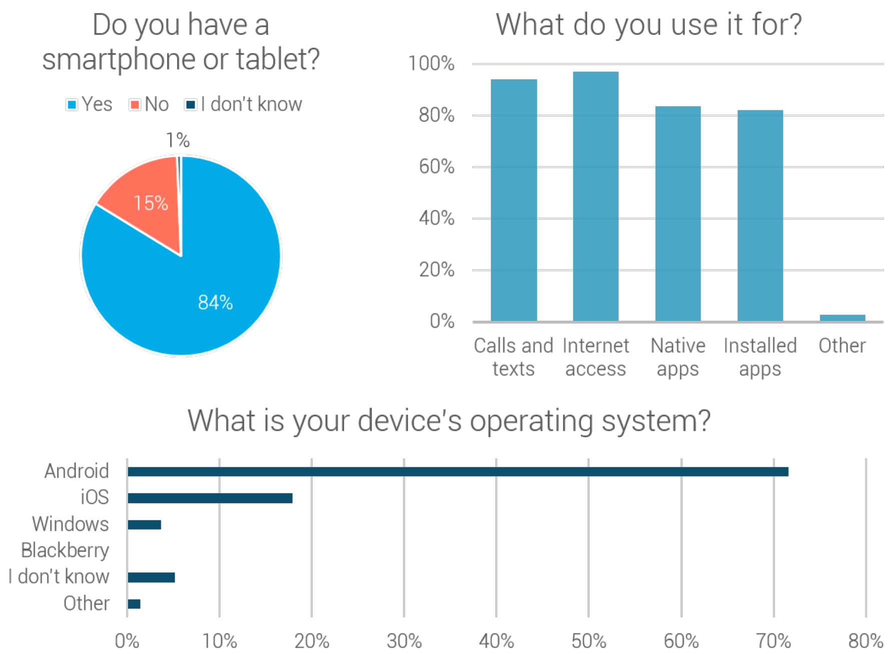


Figure 5.14: Graphs regarding the ownership and usage of mobile devices, as answered in the printed survey.

Finally, when it comes to HealthTalks, it seems most people find value in giving definitions and taking notes, but the options to record and transcribe an appointment are not as wanted. The reason for that is unclear. Even so, more than 90% of respondents say they would use the app or at least consider it, as can be seen in Figure 5.15.

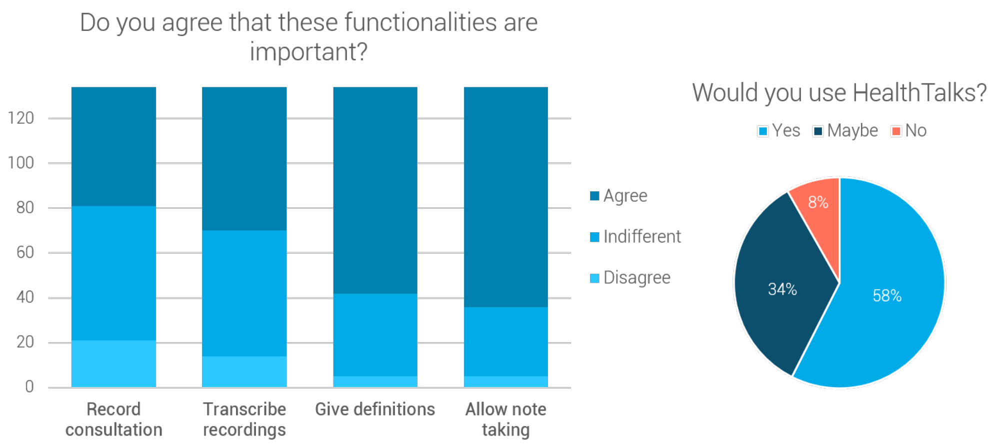


Figure 5.15: Graphs regarding the appraisal of HealthTalks' features, as answered in the printed survey.

5.3.3 Gender Analysis

As with the online survey, we had some interest in studying how the different answers related to the demographic characteristics of the respondents. To that effect, we applied the same methodology as before and what follows in Table 5.6 is the result of a Chi-squared test for $p < 0.05$. That means that these are the correlations for which there is at least a 95% certainty that the variables are not independent of one another.

Table 5.6: Association between gender and other answers in the printed survey.

Answer	Male	Female	χ^2	p-value
I take health notes and register the information of a family member	0.46	0.66	3.99	4.59e−2
I take health notes and register symptoms	0.22	0.52	9.35	2.23e−3
I take health notes and register topics to discuss in a future appointment	0.32	0.53	4.35	3.70e−2
I take health notes and I register them in a registry given by the health institution	0.08	0.25	4.35	3.69e−2
I do not always ask the physician to repeat what they said because I trust them	0.30	0.09	4.27	3.88e−2
The physician often gives me information on paper	0.30	0.12	6.15	1.32e−2

After analysing the table, it seems that women are more prone to taking health notes of a family member, and in doing so, registering symptoms and topics to discuss in a future appointment. It is uncertain why that happens, but in regards to why women are more likely to register their health notes in a registry given by a health institution, we consider that it might be because pregnant women are given such a document and men are not, so the former have access to more registry types than the latter.

Men seem more likely to receive a paper with information from their physician, while at the same time trusting them to the point of not asking them to repeat something they misunderstood. This suggests that men might have a better relationship with their doctors, though we cannot be sure from this data.

5.3.4 Age Analysis

Due to the fact that most contingency tables calculated for the age analysis had more than 20% of their cells with a count below 5, the Chi-squared test was not a reliable tool to study the independence between variables. Because of that, we decided to resort to Fisher's exact test, which is better suited to study smaller samples. In Table 5.7 below, for each age bracket we refer the difference between the obtained and expected values, and the p<0.05 for each case.

Table 5.7: Association between age and other answers in the printed survey.

Answer	18-25	26-35	36-49	50-64	65+	p-value
I take health notes and register calories consumed or spent	2.36	3.25	-4.24	-1.25	-1.13	2.90e−3
I take health notes and register blood pressure	-1.55	-4.10	0.30	-0.40	5.79	2.98e−3
I take notes of an appointment and register medication	-1.58	0	-0.70	-2.60	4.94	1.42e−2
I do not ask the doctor to repeat something I did not understand because someone will explain me later	3.11	1.69	-4.02	0.87	-1.65	1.03e−3
I have a mobile device and I use installed apps on it	2.20	5.80	1.10	-4.10	-5.03	1.40e−5

Unlike with the online questionnaire, in CHSJ we were able to get a considerable number of respondents aged 65 and older, which gives us a new perspective on the correlation between age and other variables, seeing as now we can include senior citizens in our comparisons. In fact, they appear to be more prone to register blood pressure and medication, while at the same time being much less likely to have installed apps on their mobile device. All of these characteristics appear to align with our expectations of senior citizens, who might need to care more for their health and not be as technologically inclined as younger people.

When it comes to registering calories, the younger age brackets until 35 years old appear to be somewhat more concerned than what would be expected, whereas adults from 36 to 49 years old seem to care less about that aspect than anticipated. That same group may also have less need for contacts to whom they can ask for health information, as they do not answer that they do so as often as the other age groups.

Finally, when it comes to installing apps on mobile devices, the bracket from 26 to 35 years old appears to do it more often, as do the 18-25 year-olds, albeit to a lesser degree. It makes sense that the younger generations are more technologically inclined, and that may be the justification for this disparity.

5.3.5 Qualifications Analysis

When analysing the correlation between the respondent's qualifications and their other answers we disregarded the qualifications "less than 4th grade" and "doctoral degree", since there was only one person attributed to each. The rest of the analysis followed the same methodology as the aforementioned age analysis, where we used Fisher's exact test to determine by how much each answer deviated from the expected result in each qualification group. This happened because the cell count was small for most answers and the Chi-squared test is not a good approximation under those circumstances. The results can be found in Table 5.8.

Table 5.8: Association between qualifications and other answers in the printed survey.

Answer	4	9	12	B	M	p-value
I take health notes and register calories consumed or spent	-1.51	-0.18	-3.58	2.95	2.32	4.27e−2
I take health notes and register points to discuss with the doctor in a future appointment	-3.21	-1.27	-1.20	0	3.33	3.91e−2
I take health notes and register results of exams or analysis	-0.46	-3.47	4.70	-2.90	2.05	1.63e−2
I take health notes and write them down in a note-taking app	-1.01	-0.79	-2.72	0.64	3.88	6.54e−3
I at least occasionally do not understand the physician, and it is due to their use of technical or medical terms	0.52	-1.29	4.70	-0.80	-3.10	2.37e−2
I at least occasionally do not understand the physician, and it is due to their lack of time	-2.32	-0.61	-3.80	2.62	4.10	5.83e−3
I have a mobile device and I use installed apps on it	-0.95	-3.07	5.10	-3.60	2.50	2.58e−3
I have a mobile device and I use it to access the Internet	-0.86	-1.70	1.60	0.70	0.30	2.43e−3

NOTE: The columns “4”, “9”, “12”, “B”, and “M” stand respectively for “4^{th} grade”, “9^{th} grade”, “12^{th} grade”, “bachelor’s degree”, and “master’s degree”.

We can see that people with the 4^{th} grade are the least likely to register points to discuss with their doctors, while respondents with a master’s degree are the most likely. People with the 12^{th} grade are the most likely to not understand their physicians due to the language they use; people with master’s degrees are the most likely to not understand them due to their lack of time.

5.3.6 Suggestions Analysis

Just as with the online version (see Section 5.2.6), we analysed the answers given to the question of what other functionalities could be implemented in HealthTalks. We received 41 answers (more than 1 answer for each 4 respondents), with 22% of them being invalid responses. The topics we used to classify the answers were the same as we had used previously; this was done so that we could directly compare both data sets. The results are in Table 5.9.

Table 5.9: Topics mentioned in open-ended answers in the printed questionnaire.

Topic	Explanation	Prevalence
Integration with existing systems	Having the possibility of retrieving a patient's medical history, being integrated with the National Health Service, and scheduling appointments.	27%
Reminders	Addition of a reminder or alarm system to prompt the user to take medication, undergo treatments, or go to appointments and exams.	20%
Existing functionalities	Suggestions or reinforcements of an idea that was already presented as being core to HealthTalks (recording, transcription, PHIM, or medical terms definitions).	20%
Other functionalities	Suggestions that did not fit any topic. See below for examples.	12%
Glossary	A persistent and searchable glossary of medical terms, medications, diseases, and any other relevant health information.	10%
Communication with physician	A way of communicating with their doctor or a different one to be diagnosed or receive recommendations.	7%
Health monitoring	Functionality to track the evolution of certain aspects over time (weight, exam results, medication intake, menstrual cycles), possibly in a graphic.	7%
Simplicity	The app has to be easy to use, especially by the elderly and uneducated patients.	2%

Among the "other functionalities" mentioned by respondents, we found ideas also proposed in the online survey such as the ability to append photos and prescriptions to appointments, and a map with nearby pharmacies and health establishments. We also received suggestions to expand HealthTalks to other similar fields (such as sports and nutrition), to give price comparisons between different drugs, and to provide schedules of health establishments and pharmacies.

5.4 Discussion

We decided to distribute the questionnaire in CHSJ to have a more diverse respondent demographics and, hence, be potentially more representative of the general population. We believe that the results, and specifically the characterisation of the population, corroborate our beliefs. Both surveys reveal a majority of female respondents, while complementing each other in relation to age brackets and qualifications: whereas in the online version there were very few people aged 50 years old or older, the printed version shows a more balanced representation of people from all ages. Similarly, there was an over-representation of respondents with higher education degrees in the online questionnaire, specifically bachelor's and master's (55% of the total answers), but its printed counterpart has a more balanced distribution. Therefore, we can assume that an important part of the population would not be represented in the final study if not for the printed questionnaire.

However, many answers given by both groups were surprisingly similar, with only a handful of answers being considerably different, such as the amount of people taking notes of health information. In general, there was a tendency for men to be more trusting of physicians in both surveys. The causes for that are uncertain, but a likely cause could be behavioural differences between the genders. Also, the vast majority of the respondents have mobile devices (94% in the online version and 84% in the printed version), and in both cases the majority of owners possess and Android device. Furthermore, it is interesting to note that 84% of the online respondents and 92% of the respondents in the printed version assert that they would surely or possibly use HealthTalks as it is presented. Only around 10% of the answers were negative in that question, which indicates that the development of this app is very pertinent to these populations.

On the suggestions given by the respondents, some really interesting comparisons can be drawn between both questionnaires. Firstly, the answers of the printed survey did not cover as many topics as the online version, but that can be attributed to the difference in the number of responses. The option to set reminders is very wanted by both sets of patients, but the integration with existing systems, which in the online questionnaire amounted to 7% of the responses, was mentioned by 27% of the respondents in the printed questionnaires. Our interpretation is that they were more aware of the failures and shortcomings of the National Health Service and general customer service in the health establishments because they were answering the survey while waiting for an appointment. Being confronted with or reminded of problems with the health institutions, they might have presented a bias towards that subject.

Another interesting result of the suggestions was the option to include a glossary. Even though most people who do not understand their doctors blame the language that they use, only 8% of respondents in the online version and 10% in the printed version think that the implementation of a glossary could be a good idea. This may have happened due to the lack of interest or motivation to leave feedback; most people did not leave any suggestions. Another reason could be that some people may have seen how definitions of medical concepts were included in HealthTalks' features, and did not believe that a more expansive glossary was necessary.

One thing to keep in mind is that the social desirability bias, or the propensity of answering a question with the option that is considered more socially acceptable instead of answering it truthfully [Gri10], might have altered the results of the questionnaire. That could be true for questions 3.1 and 3.4 in particular, which are respectively "Do you understand what your physician says?" and "Why do you not ask the physician to repeat what they said when you do not understand?". We tried to minimise the impact of that bias in the latter by including options that could be regarded as less usual, such as "I do not have interest" or "I lack confidence". Still, the social desirability bias may have impacted the answers to those and other questions. However, we hope that with a sample of this size such biases can be somewhat dismissed.

Chapter 6

Requirement Analysis

As discussed in Chapters 2, 3 and 4, and studied in-depth in Chapter 5, the patients' level of health literacy and their relationship with their doctor could be improved. Even though governments and independent organisations are trying to study those aspects in order to promote a better healthcare [Red17; EÁM15; HUR+15], their work may not be enough to completely revert the situation that exists in Portugal and other countries. But developers and specialists have space to elicit change in behaviours and attitudes and, hopefully, bring a quicker end to those problems.

6.1 Solution

HealthTalks aims at tackling some of the problems previously outlined by empowering the patients with tools that may ease their day-to-day health tasks and self-care ability. It is a mobile app with 4 major functionalities:

1. It can record the doctor's speech.
2. It links each recording to a customisable appointment page.
3. For each appointment, it transcribes the recording.
4. For each transcription, the app detects the medical terms used and gives small definitions for each.

The following sections expand on those functionalities by listing the requirements, personas, and user stories that guided the app's development. For more details on its features, see Section 7.1.

6.2 Product Requirements

In this section we will list HealthTalks' requirements, which were created in order to stipulate unambiguous goals and functionalities to guide its development process. Usually, requirements are the result of discussions between the customer (entity who requests the solution) and the product owner (a member of the development team) [SB02]; however, since those people overlap in this project, that process was not applicable. Our solution was to position ourselves as a user in order to better understand their needs and desires; that same process lead us to document archetypes of users, which can be seen in Section 6.3. We also consulted the questionnaire's answers (see Chapter 5) to gauge the expectations and requests of the patients, especially regarding data privacy and ethical issues. To guarantee the most thoroughness in the final result, we also investigated other works with product requirements of their own to understand what specifications ours lacked.

We started by listing the functionalities that we wished to implement in HealthTalks, seen in Table 6.1.

Table 6.1: Functional requirements.

ID	Description
FR01	The app must be able to record the audio of a conversation.
FR02	There has to be a transcription option included.
FR03	The meanings of medical concepts found in the transcriptions have to be given.
FR04	The user has to be given a way to search their appointment records.
FR05	A user should be able to take notes in the app.

Then we considered the needs of both the user and the physician, who are our stakeholders since they both interact or are affected by the app. Their requirements are listed in Table 6.2.

Table 6.2: Stakeholder requirements.

ID	Description
SR01	The recording act has to be clear and unambiguous so that both parties are aware that they are being recorded.
SR02	The privacy of the user's data must be assured.
SR03	The user must know their responsibilities when using the app.

Finally, we identified all of the restrictions that should be made for usability or efficiency purposes, as well as other quality criteria. The results can be seen in Table 6.3.

Table 6.3: Non-functional requirements.

ID	Description
NR01	The app has to be simple and intuitive to use.
NR02	A user should not have to click more than twice to start a recording.
NR03	A user should not have to click more than once to stop an on-going recording.
NR04	The array of functionalities available should be explained to the user.
NR05	The app has to be compatible with a large amount of devices.
NR06	The interaction with the chosen speech-to-text software happens through its API.
NR07	The information given to the patient has to be truthful and up-to-date.
NR08	The error rate for the transcriptions has to be 41.5% or lower.[1]
NR09	The actions of recording, checking definitions, and managing appointments have to work offline.
NR10	The app should be free to use.

6.3 Personas

According to Cooper et al. [CRC07], a persona is a design tool that represents a user archetype in order to convey how real users behave and feel. They claim that personas are the strongest models for interaction design and allow developers to focus on specific features for certain users.

They argue that there are six different kinds of personas: primary, secondary, supplemental, customer, served, and negative. Primary personas reflect the product's main functionalities and are the ones on which the development team should be most focused. Secondary personas use extra functionalities that, while useful, do not significantly alter the final solution. Supplemental personas refer only to functionalities mentioned in both previous types; their presence originates from a need to differentiate different types of users that may use the same functionalities. Customer personas only exist for products for which the person who acquires it is not the same person that uses it (such as when buying children toys). Served personas represent people that, even though they do not use or interact with the product's interface, are affected by its use (e.g. a patient doing an x-ray). Lastly, negative personas are the ones specifically not targeted by the product, but that could be mistakenly assumed as such (like early adopters of consumer products).

Personas are commonly the result of research and studies; however, since such an effort was not feasible during the conceptualisation phase, we tried to brainstorm all of the possible types of users and group them into compelling models. Five personas were developed in order to portray those archetypes; their descriptions can be found in Sections 6.3.1 to 6.3.5.

[1] Based on the best result achieved by a generic STT software when interpreting medical speech in a study by Liu et al. [LTH+11] analysed in Section 4.1.2.

6.3.1 Primary Persona: Margarida Fonseca

Figure 6.1: Photo chosen to represent Margarida Fonseca.

Age: 74
Technological proficiency: Low
Quote: "The older we get the more we need our memory not to fail us because of all the health issues we have to keep track of, and yet that's when it fails us the most."
Description: Margarida lives at home with her husband. They both have a few age-related health problems like poor sight, memory issues and frail bones, but nothing too serious or life-threatening. Those issues do require her to go to her general practitioner regularly though, and he usually redirects her to specialists according to the illness. Margarida feels blessed for having such good health for her age, but finds the amount of medical information she has to handle in her daily life very daunting and she is struggling to cope with it. She and her husband cannot rely too much on one another either since taking care of their own health is already difficult as it is.
Goals, motivations and context scenarios: See Table 6.4.

Table 6.4: Goals, motivations and context scenarios of persona Margarida Fonseca.

Goal	Motivation	Context Scenario
Remembering all her medication and treatments	Remaining independent	She takes notes of her treatment and schedules future appointments
Keeping track of her health situation	Maintaining a good quality of life	She checks the list of appointments she had for each speciality and their transcripts, as well as her personal notes for each of them

6.3.2 Primary Persona: Gustavo Brandão

Figure 6.2: Photo chosen to represent Gustavo Brandão and his daughter Ana.

Age: 29
Technological proficiency: High
Quote: “I want to give my children the best life I possibly can, but sometimes it’s hard to manage everything all by myself.”
Description: Gustavo is the proud father of two small children. He does mostly odd jobs, which means he has more time for taking care of them than his girlfriend does, since she has a full-time job. Their eldest child, Ana, has Crohn’s disease, which requires special attention from them. Gustavo has to manage her doctor appointments, as well as her medication and treatments, not to mention their son Hugo’s healthcare, and his own. He gets by thanks to a small journal he carries everywhere and where he writes down the doctors’ orders and any other relevant information. The problem is that he sometimes does not fully understand what he is told and is too intimidated to ask. When that happens he writes what he hears and when he gets home he tries to search for those terms online to learn more about what they mean.
Goals, motivations and context scenarios: See Table 6.5.

Table 6.5: Goals, motivations and context scenarios of persona Gustavo Brandão.

Goal	Motivation	Context Scenario
Organise each person’s health information	Get better and faster access to the details of each family member’s healthcare situation	He associates each appointment with the family member in question to be easier to find
Understand the medical terms	Get a better grasp of his children’s health status and how to improve it	He transcribes the medical appointments and checks the definitions of the terms he does not know

6.3.3 Supplemental Persona: Diana Magalhães

Figure 6.3: Photo chosen to represent Diana Magalhães.

Age: 18
Technological proficiency: Very high
Quote: "Even though my condition does not define me, it never really leaves me. So I just have to work around it in order to be as efficient as anyone else."
Description: Diana has a hearing impairment, but you would not notice if you did not know. She is young, but since she has never been able to hear, she developed her own means to circumvent that, mostly based on lip-reading and contextual clues. However, now that she is starting to be her own carer after reaching adulthood, she is noticing a stark difference between everyday language and medical speech. The latter is much more difficult for her to understand, since it contains a lot of words that are mostly unknown to her. She often asks her physicians to communicate with her by writing, but that can quickly become cumbersome to both.
Goals, motivations and context scenarios: See Table 6.6.

Table 6.6: Goals, motivations and context scenarios of persona Diana Magalhães.

Goal	Motivation	Context Scenario
Understand the physician's speech	Be more independent and less inconvenient	She records her physician and transcribes their speech so she can read what they said
Understand the medical terms	Learn new words and their enunciation	She checks the transcription for medical terms and sees their definitions

6.3.4 Supplemental Persona: Luís Correia

Figure 6.4: Photo chosen to represent Luís Correia.

Age: 46
Technological proficiency: Medium
Quote: "I am used to taking notes on paper, but I don't like to manually look for each patient's notes on all my notebooks each time I want to check something."
Description: Luís is a psychologist. His profession demands that he always remembers the details pertaining to each patient's life every time they visit him, so that they feel heard and the conversation can flow easily every time. In order to brush up on each individual's life and behaviours, he reviews their processes the day before the appointment. However, since he takes notes on paper, he finds himself often piecing together information from different notebooks or separate papers.
Goals, motivations and context scenarios: See Table 6.7.

Table 6.7: Goals, motivations and context scenarios of persona Luís Correia.

Goal	Motivation	Context Scenario
Record the main points of each patient's accounts	Remember them in later appointments	He records himself recapping the session he had with a patient and associates the patient with the appointment
Get faster access to each patient's history	Be more efficient	He searches for all of the appointments associated with a specific patient

6.3.5 Served Persona: Rute Tavares

Figure 6.5: Photo chosen to represent Rute Tavares.

Age: 32
Technological proficiency: Medium
Quote: "Even though I try to treat each patient as a unique individual, it's difficult to do so after long work hours and strict schedules."
Description: Rute is a general practitioner at a hospital and loves what she does. However, it can get quite tiresome, since the hospital is lacking doctors and so she often has to work overtime in order to make up for that. She notices that at the end of the day her concentration decreases noticeably and her actions become more automatic. That leads her to treating patients in a more detached way, not being as attentive to the degree to which they understand what she tells them, and trying to quickly wrap up medical appointments. She is aware of those flaws and tries to correct them every chance she gets, but sometimes it is very hard to perceive those faults in real time.
Goals, motivations and context scenarios: See Table 6.8.

Table 6.8: Goals, motivations and context scenarios of persona Rute Tavares.

Goal	Motivation	Context Scenario
Ensure that patients understand her	Empower them with important information	She encourages her patients to use HealthTalks and resort to it when in doubt of a term's meaning
Guarantee that her patients do not forget what she says	Avoid that they mishandle their own self-care	She impels her patients to use HealthTalks to manage past appointments and take notes

6.4 User stories

A user story describes a system functionality with informal and natural language. They are different from functional requirements because the latter focus on the product's features instead of the user's intent and motivation. Moreover, requirements may be seen as less holistic and consistent. User stories are usually used as a way of communication between the product owner and customer, and among developers [Coh04].

HealthTalks will only have one type of user, referred to simply as "user". In Table 6.9 all of HealthTalks' functionalities are listed based on the personas designed in Section 6.3.

Table 6.9: User stories.

ID	Description
US01	As a user, I want to record the audio of a medical appointment in order to check the physician's recommendations in the future.
US02	As a user, I want to read the transcribed text of the audio of a medical appointment in order to skim through it quicker.
US03	As a user, I want to see the definitions for the terms that I don't know in the transcription in order to acquire new knowledge of my health condition.
US04	As a user, I want to be able to edit all the details pertaining to a medical appointment in order to make it as detailed as possible for future reference.
US05	As a user, I want to be able to search for appointments by their different fields in order to find the one I'm looking for faster.

Chapter 7

HealthTalks

In this chapter, we will focus on HealthTalks and all its components: why and how they were built, how they interact with one other, and describe the tests they were submitted to. In the end, we present a broad overview of the system as a whole, its limitations and potential improvements.

7.1 System Description

As already mentioned, HealthTalks' ultimate goal is to improve the patients' capabilities for a proper health management. In order to achieve that objective, it first records the dialogue during a medical appointment (the interface for this and the other features can be seen in Appendix C). The purpose of this feature is to give patients a way to recall what was said by the doctor (such as treatment details, medication dosage, and other pertinent information). That recording should happen if both parties consensually agree to do so, which is why we implemented an always-on display while recording, which stops recording when turned off and exhibits a very bright microphone icon and a loading animation as a way of disclosing its purpose.

However, the appointment does not need to be fully recorded for the patient to benefit from this functionality. Arguably, the best solution would be to record a short excerpt of the physician's advice at the end of an appointment, including all of the important talking points. In order to allow the patients to check that information faster and more easily, we implemented a transcription system that converts the audio to written text. This feature adds convenience and clarity to all users, but can also improve the app's accessibility by patients with special needs, such as those with hearing impairments.

Even after the doctor's speech is recorded and transcribed, the patient may still have some difficulties understanding their meaning. As seen in Chapters 3 and 5, many people do not understand their doctor at least occasionally, and often due to their wording. Because of that, we implemented a way of detecting medical concepts in the transcription and providing short definitions for them. If those definitions prove ineffective, we also give the option to learn more about each individual issue online, which may contribute to more proactive patients.

All of those functionalities are associated to a medical appointment, but the app was built to support several appointments. That is why each of them also allows for the introduction of specific fields such as title, date, patient (in case the user has dependants for whom they wish to record appointments), physician, medical speciality, and health institution. An appointment may also be signalled as important to be further differentiated. Furthermore, there is also a section that allows the user to take notes about the medical appointment in case they want to register any other information. These fields allow for a more precise search function that allows the patients to quickly find the information they want. The fact that the appointment's date can be edited allows for patients to plan ahead for future medical appointments as well.

The main novelty in HealthTalks is holding the patient accountable for their own health condition while supporting their methods of doing so. It recognises weaknesses, such as low health literacy levels and bad personal health information management practices, and gives patients a tool to overcome them specifically in the context of medical appointments. A comparison between HealthTalks and some of the existing apps, analysed in Section 4.2, can be seen below in Table 7.1.

Table 7.1: Apps similar to HealthTalks and the ways the latter differentiates itself from the competition.

App type	Examples	Differentiation
Health information management	MyFitnessPal, MyChart, BG Monitor Diabetes, Pillboxie	HealthTalks is all-encompassing, being relevant for any type of health issue
STT note-taking	Evernote, ListNote, Speech-Notes, Google Keyboard	HealthTalks translates Portuguese texts and provides definitions for the technical terms
Health information	Patient Journey App, Patient.co.uk, eMed.pt, MySNS	HealthTalks has less extensive information, but provides definitions for the concepts most relevant to the user's health condition

The inclusion of target users in steps of the development (e.g.: survey to understand their difficulties, habits and needs; user study to assess the app's interface) makes us believe that HealthTalks is suitable to its audience. All of its parts were created thoughtfully and purposely to give the best possible user experience, while being unambiguous with its goals.

In the next sections we will focus on the two external services that we use regarding speech recognition and the retrieval of health vocabulary.

7.1.1 Speech Recognition

In order to select a speech recognition software that was appropriate for our implementation, we developed a process to compare and test them.

The first step was to look for the most used speech-to-text (STT) solutions in the market and compare them. A summary table can be seen in Table 7.2. We focused on objective variables such as price, language options, and personalised dictionaries, as well as more subjective characteristics among which were the amount of documentation available, reviews from other users, and ease of integration with Android. Another option to think about for future iterations of the app would be the ability to use the STT system offline, which some solutions provide.

Table 7.2: Comparison between different speech-to-text options.

Name	Languages	Price	Customisable
Android Speech-to-Text[1]	EN, PT	Free	Yes
Bing Speech API[2]	EN, PT	5000 free transactions	–
DynaSpeak[3]	EN	1000 minutes	–
Dragon NaturallySpeaking 13[4]	EN	169 Euro	–
Google Cloud Speech API[5]	EN, PT	Free 1000 mins/month	Yes
HTML5 Web Speech API[6]	EN, PT	Free	Yes
IBM Watson Speech to Text[7]	EN	Free 100 mins/month	Yes
iSpeech[8]	EN	Free	Yes

"EN" stands for English and "PT" for Portuguese; "mins/month" is minutes per month.

We immediately disregarded any options that could not be used for free, as well as all of those that did not have the option of transcribing recordings in European Portuguese. Since the remainder of the options were relatively similar in terms of their characteristics, we chose to rely on our impression on the amount of support and examples each provided and so, from those options, we decided we would consider Bing Speech API and Google Cloud Speech API, since they are both free, support European Portuguese, and are supported by major technological companies. To select the one more adjusted to our needs, we decided to test them and compare the results. An in-depth explanation of those tests can be found in Section 8.1.

After those tests were complete, we decided to use Google Cloud Speech API (GCSA) due to its superior results. The GCSA has three different ways of delivering speech recognition: it can

[2]https://developer.android.com/reference/android/speech/package-summary.html
[3]https://docs.microsoft.com/pt-pt/azure/cognitive-services/speech/home
[4]https://dev-portal.api.sri.com/#/dynaspeak
[5]http://shop.nuance.co.uk/store/nuanceeu/en_GB/pd/productID.305733200/pgm.95401100/OfferID.12345678910/Currency.EUR
[6]https://cloud.google.com/speech/
[7]https://dvcs.w3.org/hg/speech-api/raw-file/tip/speechapi.html
[8]https://www.ibm.com/watson/developercloud/speech-to-text.html
[9]https://www.ispeech.org/developers/android

be either synchronous recognition, during which an audio sample of 1 minute or less can be sent to the API and there will be a wait for the response; asynchronous recognition, which supports larger audio files up to 80 minutes in duration and during which the program can periodically check for results; and streaming recognition, in which the recognition is done over a bi-directional stream with interim results being sent and changed while the user is speaking. Among all of these options, the asynchronous recognition is the most applicable in HealthTalks, since we expect that the recordings will be longer than 1 minute, and the streaming would require Internet usage while at a medical appointment, which is something we wanted to avoid in order to include people with little to no mobile data service.

At the same time, the API supports communication through REST (representational state transfer) or gRPC, an open-source remote procedure call (RPC) framework. One advantage of gRPC over REST is its faster speed (it can have up to 25 times higher performance than REST [Hus16]); and the fact that it uses protocol buffers also allows for some significant differences. Protocol buffers are a way of structuring serialised data, just like XML or JSON, that allows for a higher level of abstraction in their implementation. They are more condensed than their equivalents, which means that they tend to be smaller in size (up to 10 times smaller than XML [Goo17b]); and they are also not intended to be read by humans, unlike JSON for example, which means that if they are intercepted, they will most likely be impossible to interpret without access to the schema used for their encoding [Ogi16]. This brings several advantages to the implementation, including the reduction of Internet usage for the user due to smaller packages being sent and received, and a barrier to potential data interceptions that increases the privacy of their information. Moreover, the streaming speech recognition option offered by the GCSA is only possible through gRPC, so if we use that type of communication we allow for an easier expansion of the app in the future, in case such a feature is ever needed.

Another feature of GCSA is its ability to accept a list of phrases or terms alongside an audio file. Even though GCSA already has a vocabulary of its own that it uses to recognise the words in the audio, its transcriptions work by measuring word probabilities. This means that if a word is very unlikely to be said either at all, or in a specific order relative to another word, another term with similar phonetics may be chosen for the translation instead. To counteract that, the developer may use the phrases feature to indicate a list of terms that are more likely to be said in the given audio file than what would be usual in a common conversation, in order for GCSA to prioritise those concepts over other similar ones. In our project, this feature gains relevance since the medical field includes a lot of very specific vocabulary that may be uncommonly used, such as the names of several disorders or anatomy concepts. However, Google limits the amount of phrases a request can send. The limit is 500 phrases per request, with a total of 10000 characters or less, and each phrase having 100 or less characters. This is somewhat limiting when considering the amount of available medical concepts; for example, the number of articles under the Portuguese Wikipedia's category “Doenças” (diseases and disorders) and its subcategories is 1400. We still tested this functionality in a specific setting (see Section 8.3), but decided not to implement it due to the poor test results and the aforementioned limitations on the amount of phrases.

Ultimately, we chose to implement the GCSA through gRPC by employing an Android example created by Google[10] and adjusting it to our specific needs. When it came to its implementation in the app, we created a confirmation window before employing the transcription capabilities in order to check if the user was connected to the Internet through mobile data; if so, a warning appears to make sure that the user understands that this task will spend data from their plan. The final transcriptions and the use of phrases to obtain better results were evaluated in Section 8.3.

7.1.2 Health Vocabulary

In order to give definitions of the medical terms found in transcriptions, a glossary was needed, so it was crucial to find a database of medical concepts to use. Our restrictions were that it had to be inclusive of all medical specialisations, since each patient might have different needs; and that it had to have the option to be storable on the device for offline use. This last requirement lied within our wish to make the app accessible for everybody: since some people have limited or no mobile data plans on their smartphones, we wanted to make sure that they were still able to get definitions in places with no wireless networks (such as in many health establishments). That, in turn, also restricts the size of the database due to storage limitations.

Even so, the speech-to-text process still needs an Internet connection, and the definitions are assigned after that process. However, our intention was to keep our possibilities open for potential feature additions that could require the use of the glossary, such as a searchable dictionary of medical terms. That was a very requested feature among respondents of the questionnaire (see Chapter 5); by not needing an Internet connection to acquire the glossary terms, we allow for that possibility in the future. Another aspect to take into account in order to allow for more flexibility in future implementations is personalisation. Different patients with different health conditions will probably be confronted with different vocabulary in their medical appointments. Furthermore, some people might want to use the app in slightly different contexts such as veterinary visits or dentist appointments. For those cases, it would be a benefit if the glossary could eventually be changed to be adaptable and customisable by the user.

To allow for those kinds of possible iterations, our possibilities were extremely narrow. Even without such restrictions, most existing Portuguese databases are translations from medical authorities, like the WHOPOR (Portuguese translation of the WHO Adverse Drug Reaction Terminology), the MDRPOR (Portuguese translation of the Medical Dictionary for Regulatory Activities), and the ICPCPOR (Portuguese translation of the International Classification of Primary Care). This means they may have a high degree of specificity and their definitions are not always easy to understand [Sou16].

Our final choice was MediaWiki API[11] (MWA), a RESTful web service that allows access to Wikipedia's content, both for queries and editing. Since Wikipedia's language is usually colloquial, its articles are categorised, and it is regularly updated, it made sense to use it as a data source for the definitions. Moreover, the MWA has a lot of flexibility, allowing for requests in

[10] https://github.com/GoogleCloudPlatform/android-docs-samples
[11] https://www.mediawiki.org/wiki/API:Main_page

different languages, different types of results (such as categories, articles, and discussions), and a high degree of customisation through the use of several variables (such as limiting the articles in the response by their last edition date, or changing the response format of the API request).

On the veracity of Wikipedia articles, several studies deem them considerably truthful. Research has shown that Wikipedia is almost as accurate as *Encyclopaedia Britannica* [BBC05; RMM+12], with an accuracy rating of around 80% (with published encyclopaedias presenting rates of around 95%) [Hol08], though the structuring and phrasing of its articles is often pointed as a weak point. In regards to health information, Wikipedia is referenced as an accurate and comprehensive source on mental health topics [RMM+12] and pharmacology [KMG+14], although some experts recommend the consultation of other information sources to complement Wikipedia's information. As for the update rates, a study on Wikipedia articles for drug prescriptions found that 41% were updated within two weeks of the most recent FDA (USA Food and Drug Administration) safety warning; others were updated later, and 36% were not updated for at least one year after the warning [HBS14]. Even though we do not have statistics for the Portuguese Wikipedia, we believe the results should be similar, and that it is a good starting point for a patient to learn more about a concept they do not understand.

The nature of Wikipedia's article categorisation is somewhat chaotic with the intention of creating as many logical links between articles as possible. Its guidelines[12] detail that every article has to belong to at least one category and usually they cannot belong to both a category and one of its subcategories. However, there are exceptions to that rule, such as non-diffusing subcategories, which means that the overall structure is not organised in a tree or a directed acyclic graph, allowing for situations such as a category being a descendant of itself [KCS09].

With that in mind, we cannot continually look for subcategories or we may end in an infinite loop. A first iteration of the query was done to retrieve articles with depth 3 in relation to the category "Medicina" (medicine) in the Portuguese Wikipedia. That depth meant that it would retrieve all of the articles immediately descendant of the main category, as well as the ones under all its subcategories, under the categories below those, and finally under the latter. However, the query was very slow, with an average of approximately 35 results per second. In order to increase the query speed, the PetScan[13] tool was used, which can send requests to any Wiki with a faster speed. Using PetScan, the average return was of around 3216 results per second, or around 92 times more results in the same amount of time.

However, the disordered nature of Wikipedia's categories means that going only as far as depth 3 when looking for medicine pages, as in the previous example, we get 14582 results. Among those, as an example, we can find an article about Pyotr Chardynin, a XIX century Russian film director, screenwriter, and actor; and the page for the "Got Milk?" American advertising campaign. Because of that, we found it more useful to restrict the depth to 1 (meaning only one level of subcategories will be checked), while setting "negative" categories found among those that we do not wish to include (such as ones about historical figures or awards). The current query is

[12] https://en.wikipedia.org/wiki/Wikipedia:Categorization
[13] https://petscan.wmflabs.org/

then "medicine" combined with "medical terms", "anatomy", and "diseases", excluding the categories "medicine history", "medicine awards", "medicine in fiction", "obsolete medical terms", "medicine journals", "disease survivors", "Brasil medicine", "deaths by disease", "eradicated illnesses", and "animal diseases", with a depth of 1, and only including articles updated since 2015. For each result, we query the first 3 sentences of the article. This returns 2923 results and requires around 680KB of storage. Further testing would be necessary to confirm if those terms would all be useful in a typical medical appointment, or if some important definitions are missing, but the query can be easily changed at any time.

That query is executed only when the user first opens the app, and on subsequent updates to the app. This means that all of that information can be checked in real-time by the user, even if offline, while never being outdated as long as the app is updated somewhat regularly. The user can also update that database in the settings if they so wish, and a possible option to consider in future implementations would be customising the database to the user's needs (such as adding more paediatric vocabulary if the user has children).

7.2 System Architecture

As seen in Figure 7.1, the solution can be divided in three major layers.

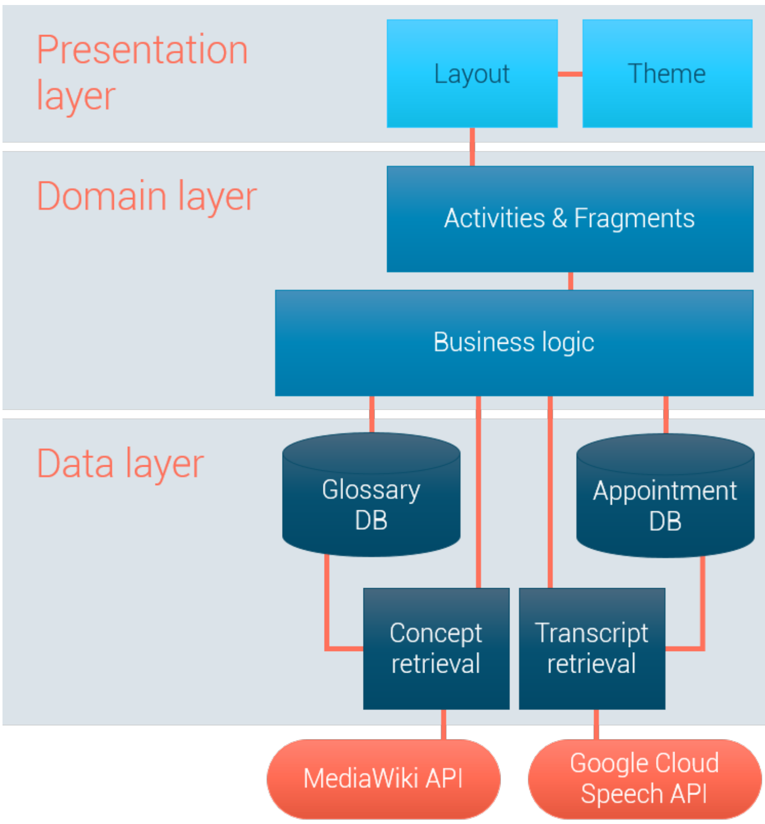


Figure 7.1: Architecture of the HealthTalks app.

- Presentation layer: responsible for drawing and positioning the interface's elements. It can be divided into "Themes", which includes all graphics, icons, logos, colours, and fonts; and "Layout", which is composed of several XML files detailing the structure of each interface. In the latter we can also include the implementation and drawing of app shortcuts, which appear after a long press on the app's icon before opening it. More information regarding the user experience can be found in Section 7.3.

- Domain layer: includes most of the app's logic functions. "Activities & Fragments" is the parcel that communicates with the presentation layer by changing interface elements in real-time, specifying on-click commands, and structuring the activities, fragments, and their relations. The "Business logic" module is responsible for features such as the voice recording and, more importantly, the coordination of all the other components.

- Data layer: here we can distinguish between the two databases (with their respective schemas and handlers), which store appointments and glossary terms to use in the transcriptions, and the two data retrieval modules: the component "Transcript retrieval", that communicates with Google Cloud Speech API in order to transcribe the voice recordings and store them in the corresponding database; and the component "Concept retrieval", that retrieves information from the MediaWiki API. The first only happens when the user requests it, and the latter when the app is opened for the first time and whenever it is updated by the developer. More information on the use and implementation of these external services can be found in Sections 7.1.1 and 7.1.2, respectively.

We chose to implement the solution in Android based on a) the ubiquity of that operating system nowadays [Nal16] (confirmed by the answers to our questionnaire in Chapter 5); b) the familiarity of the authors with its development tools; and c) the easier access by the authors to Android devices for testing. The oldest version we are supporting is Android 4.0.3 Ice Cream Sandwich, released in 2011. Other operating systems may be supported in the future.

The solution was developed mostly in Java (for the domain layer and most of the data layer), SQLite (for the database handlers and queries), and XML (for the presentation layer) in the IDE Android Studio[14]. The communication with the MediaWiki API was done through a RESTful interface in Java, with the results being given in JSON. The communication with the Google Cloud Speech API was done through gRPC in Java using protocol buffers.

7.3 User experience

The user experience (UX) was considered a major aspect of this app, mainly because of our aim to make it accessible to the general population. This includes people who might not be very proficient with the use of smartphones and people with somewhat limited capabilities. As such, the layout

[14] https://developer.android.com/studio/index.html

and journey paths were carefully built since the beginning taking into account that everything should be as simple and intuitive as possible.

According to Nielsen [Nie94], usability does not consist solely of an interface's learnability, but also its efficiency when used by proficient users, its memorability for recurring users, its enjoyability, and its low error rate. All of those aspects were taken into account when creating the app's interfaces and are studied in more detail in the following sections.

7.3.1 General Layout

The interface presents two major colours, blue and orange. This contributes to a more simplistic and minimalistic look, while at the same time having enough diversity to accomplish each view's goals. The orange colour is used sparsely throughout the app in order to have more impact and convey more importance when shown. Hence, it was used in the places that needed to catch the users' attention faster. More information on the project's graphic identity and underlying decisions can be found in Appendix D.

While everything in the main screen is blue, the recording button is differentiated by being orange and slightly bigger. It was our intention to make it stand out from the other ones and be easier to press, since it is the only option that the user will need to find with limited time during an appointment. All the others may be more similar to each other because they should not be as urgent to get to as this one. Still, since Fitt's law determines that the movement time needed for a person to click on a target is inversely proportional to the width of said target [Zha02], there was an effort to make all of the buttons relatively big in order to enable a faster experience.

In a similar vein to the aforementioned colour differences, the recording page is mostly orange in order to catch the attention of anyone looking at the smartphone: it is important to make it clear that the conversation is being recorded. The microphone symbol and the loading icon are also highlighted in that page. The appointment page was divided in tabs to make it easier for the user to find the content they are looking for.

When it comes to structural decisions, HealthTalks' layout is mostly in accordance to Google's Material Design guidelines[15], since they have become a standard among Android apps. Also, specific attention was given to devices of different sizes, from smartphones to tablets, to make sure that the layout was clear and coherent in all aspect ratios and pixel densities. Both portrait and landscape modes were considered for each device, and some examples can be seen in Figure 7.2.

7.3.2 Learnability and Memorability

As previously mentioned, Nielsen [Nie94] considers that a good user experience implies, among other aspects, good learnability and memorability. We tackled those issues firstly by creating an app that closely follows the Material Design guidelines from Google. This is an important aspect of learnability, since most Android smartphone users will likely have encountered an app with this layout before, and will therefore have a less steep learning curve with HealthTalks.

[15] Available at https://material.io/guidelines/

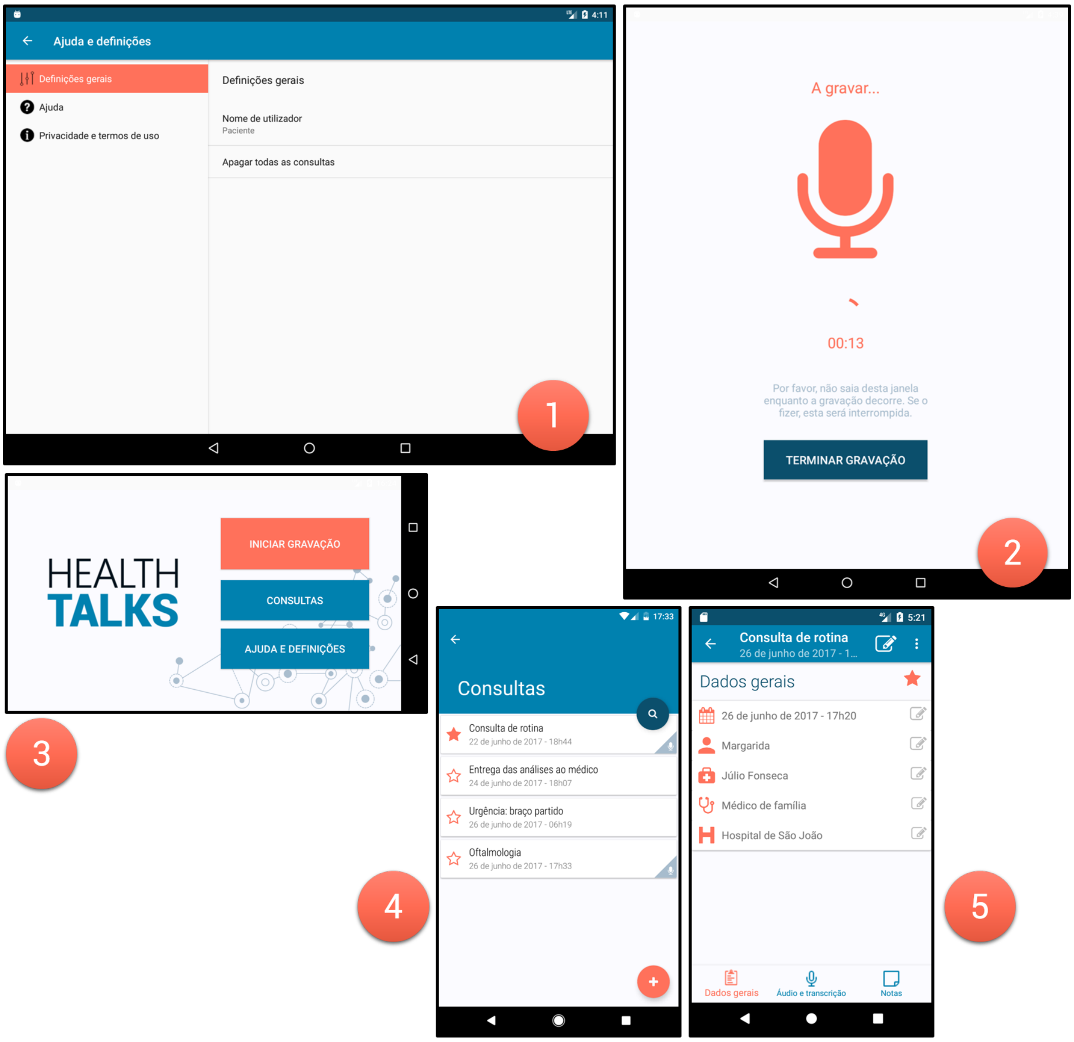


Figure 7.2: Some of HealthTalks' interfaces in different devices and modes. 1: settings view on a Nexus 7 tablet in landscape mode; 2: recording view on a Nexus 9 tablet in portrait mode; 3: home view on a Nexus 5X smartphone in landscape mode; 4: appointment listing view on a Google Pixel smartphone in portrait mode; 5: appointment view on a Galaxy Nexus smartphone in portrait mode. The screens are not to scale.

Secondly, we tried to include as many icons as possible throughout the app in order to create mental associations between the icon and the button's functionality or the concept in question. Hopefully that will lead to a higher memorability, since humans are better at memorising pictures than words [GMR+98].

7.3.3 Efficiency by Proficient Users

A utility that we implemented for the most advanced users was the recently released Android 7.1 Nougat's "App Shortcuts" feature, which allows a user to press an app's icon for longer than usual in order to reveal custom shortcuts to specific app views, set by the developer. The options chosen for HealthTalks were "Record appointment" since, as said before, it should be an easily reachable option due to the user's time limitations, and also "View appointments", which opens the appointments list. This last option was chosen so that the user could easily check their notes or past appointments. No other shortcuts were implemented to avoid overwhelming the user.

Not only that, but we also adapted our interface to proficient users by giving them more information at a glance. An example of that is the presence of a microphone icon in the lower right corner of some appointment cards, which indicates that that specific appointment has a recording associated with it. This enables a more efficient use of the app by making it unnecessary to open the appointment to check if it has a recording. Examples of both features can be seen in Figure 7.3.

Another feature that can be used by proficient users is the search engine. It can simply accept a string before pressing the "search" button (and then it will search all fields in all of the appointments), but it can also be used in different ways. One such way would be to enable the "only show favourite appointments" option and search with only that field selected. If the user does so, the app will automatically show only the favourite appointments, without any other type of filtering. That may seem counter-intuitive to some users, but we believe that the most technologically proficient ones will be able to benefit from the added complexity of search options.

Even though these implementations are aimed at more advanced users, they are unobtrusive and mostly invisible to the more common users, which leads to an effective user experience adjusted to the user's skills.

Figure 7.3: Features for more advanced users: on top, app shortcuts shown on a long press of the app's icon; on the bottom, an appointment card with the microphone icon in the corner.

7.3.4 Enjoyability and Low Error Rate

To increase a user's enjoyability of the app, we tried to help them avoid unnecessary or frustrating tasks. One way we did so was by allowing them to set the default patient name in the app's settings. This will make all existing appointments with a blank patient name to receive that new name, and any new appointments created in the future will also be set to that patient. This will help people who use the app mostly for one person to keep all of their appointments organised with less effort.

In the same vein, we implemented an automatic title creator, so that the user does not have to set the titles for each appointment. When the user specifies the doctor, health establishment, or speciality of an appointment, a title is automatically generated with some or all of that information. This title does not overwrite the existing title if the latter was created by the user, and can be edited at any time.

Lastly, the error rate of the app in general is something that will have to be thoroughly tested in the future, but in our experience it works smoothly and as intended.

7.4 Limitations

7.4.1 Ethics

There are several different ethical dilemmas to consider in a healthcare context, including clinical care, health services and systems, public health, epidemiology, bio-technology and the use of animals in research [Glo15]. However, the only one relevant for this project lies with the physician-patient confidentiality, which may be broken with HealthTalks due to the recording of medical appointments.

Confidentiality is the basis on which medical practice is built, since health professionals have access to an individual's personal health information and have the obligation of keeping it private unless consent to release the information is explicitly given by the patient or the law enforces it [BBD13].

Since this project aims at putting the information in the hands of the patient, the physician's Hippocratic Oath [Eva14] does not interfere in any way with what we are trying to accomplish. However, the collected information can still be used against the patient's will (for more on that, please see Section 7.4.2), or against the physician's privacy.

In the USA, health professionals are often the target of prosecution and there is a very noticeable position from government officials against any possible errors from their part, being that health fraud was the highest priority "white-collar" crime, even above violent street crime, after the September 11th 2001 terrorist attack [Lib08]. Therefore, many physicians may feel reticent towards an app that records their every word during an appointment due to fear of backlash from the patients.

However, two aspects should be acknowledged: firstly, there are several different audio recording options available to everyone at the moment, all subject to the same privacy concerns as HealthTalks in regards to recording the physician's dialogue. Many of them are Android apps, such as “Secret Voice Recorder”[16], which markets itself as a concealed recording solution, and in that way presents an even greater potential for undermining the physician-patient confidentiality. Nonetheless, HealthTalks unambiguously encapsulates a recording feature in a healthcare-oriented software, which may be misinterpreted as a reassurance of the right to covertly record a medical appointment. But secondly: measures can be taken to decrease the possibilities of patients recording

[16] Available at https://play.google.com/store/apps/details?id=com.miragestack.secret.voice.recorder

their physicians against the latter's will. One such measure, already implemented, is the constraint of being forced to have the device's screen turned on with a microphone icon displayed in a vibrant colour for the duration of the whole recording process. Other efforts that may be implemented in the future include a loud signal when starting the recording; or not allowing recordings longer than a pre-determined amount of time. This last point may even be turned into a main feature of the app: instead of recording the entire appointment, we might advocate for recording only a few minutes of the appointment, during which the patient asks the physician to repeat the most critical information they shared (such as diagnosis, medication, dosage, and treatment).

A new questionnaire is intended to be shared among health professionals after the end of this dissertation with the goal of inquiring them on their views on this subject and plan further measures to implement in the app according to the responses.

7.4.2 Privacy

The components of HealthTalks developed during this dissertation do not require Internet access nor any other potentially vulnerable external communication system. All of the data (the recordings, transcribed text, and custom fields) are saved locally in the device used. No login system was implemented either, so the safety of the data will be assured on the assumption that the device itself is safe. However, if the device is breached, all of that data will become vulnerable since it is in no way protected from someone with access to the device. Moreover, there is a third party API used remotely to transcribe the recordings, which means that those files (both the audio files and the text transcriptions) will be even more vulnerable to attacks when sent and received from the API. Even though they are being sent to a third party, the responsibility of protecting the user's privacy lies within the main application [Com17]; in this case, HealthTalks. On top of that, the user has to be informed that their data will be shared with a third party in order to develop the transcription.

According to the new European Data Protection Regulation to be implemented after May 25th 2018 [Com17], the users have the right to erase their data at any time. Moreover, the data in HealthTalks is deemed personal data since it pertains to an individual's health and biometric information [Eur16], so further security measures have to be ensured, including the manifested will of the user to share that information. Finally, a system of notification of the Portuguese Data Protection National Centre in case of safety violations by a third party has to be developed. The full extent of the knowledge that has to be given to the user is as follows (reproduced *ipsis verbis* from the 2016/680 European Directive [Eur16]):

- the identity and the contact details of the controller;
- the contact details of the data protection officer, where applicable;
- the purposes of the processing for which the personal data are intended;
- the right to lodge a complaint with a supervisory authority and the contact details of the supervisory authority;

- the existence of the right to request from the controller access to and rectification or erasure of personal data and restriction of processing of the personal data concerning the data subject.

The main feature that was added to address this issue was a privacy declaration where the user is informed, in layman's terms, of every aforementioned item. We also implemented an option to delete each appointment individually or all of them at once, clearing all of the user's data from the device. Besides that, an informative note was added with details on how to report any privacy concerns to the Portuguese Data Protection National Centre, which is the applicable protection officer.

All of these privacy measures were taken into account when developing the app. Even though it will not be freely available before the end of the dissertation, it is important to have all of those requirements in mind since some will affect how the app continues to be developed in the future.

Chapter 8

Evaluation

8.1 Testing Speech Recognition Software Options

Tests were carried out to ensure we were using the best option of STT software among the two previously selected: Bing Speech API (BSA) and Google Cloud Speech API (GCSA). To learn more about the selection process that resulted in those two options, please refer to Section 7.1.1. The tests consisted in the transcription of two short texts in English, one generalist and another with medical jargon. Our goal was to compare their performance in both situations and ideally, if the results were significantly different, choose the one that performed best to be used in HealthTalks. We did not test Portuguese texts because the online resource used to test BSA did not include it as an option.

The texts (available in Sections 8.1.1 and 8.1.2) were read by a researcher out loud in order to be processed by the programs, which posed two problems. The first is the fact that the researcher does not know English natively and does not have a proper British, American, or any other native accent. The second is that, since the texts had to be repeated for each program, the enunciation could have been slightly different in each take.

In response to the first situation, we do not think the accent interfered with the results. It may have increased the error rate in the transcriptions, but since the input in both programs had the same accent, it should have increased the error rate by the same amount in both. This means that, even if the error rates do not correspond exactly to the actual values, the difference between the programs — which is our focus — should not have changed.

The use of different takes was a necessity due to the lack of multitasking in the device we used. We could have recorded the voice, but we lacked a device with enough quality to save and play it appropriately. Still, there was an effort to minimise the impact of repeating the texts by training the enunciation several times before recording, in order to pause in the same places, read at a constant rhythm, and enunciate the words in the same way. Because of this, we do not believe it played a significant role in the results.

The tests were conducted using samples of the aforementioned software available online. The GCSA was used by installing its Android Package (APK) based on the official repository[1], while the BSA was used through Microsoft's web platform[2]. The tests were conducted in a Google Pixel smartphone, which uses Qualcomm Fluence Pro technology to enable three-microphone noise cancellation. The phone was resting on a table, about 35 centimetres from the sound source in a controlled environment with some minor and unobtrusive ambient noise. Our goal was to approximately replicate a possible use case scenario in a physician's office.

8.1.1 Generalist text

Since we wanted a basic English text with which to test the software, we looked for a medium-difficulty sample with no complex words. To achieve that, we searched writing samples around the B2 level, as measured by the Common European Framework of Reference for Languages: Learning, teaching, assessment (CEFR). That level assumes an ability to convey straightforward concepts and hold conversations with some ease, being classified as an intermediate level for foreign speakers [Cam15]. The selected sample was adapted from a "Reading and Use of English" section of a Cambridge English: First (FCE) sample paper[3] in order to have exactly 100 words:

> We live on the island of Hale. It's about two kilometres wide at its broadest point, and it's joined to the mainland by a narrow road built across the mouth of the river which separates us from the rest of the country. Most of the time you wouldn't know we're on an island because the river mouth between us and the mainland is just a vast stretch of tall grasses and brown mud. But when there's a high tide and the water rises a half a metre or so above the road, then you know it's definitely an island.

8.1.2 Medical text

We wanted the medical text to represent what a physician could realistically be expected to say to a patient. This means that we did not want a continuous enumeration of diseases and illnesses, but rather a balanced view on a specific problem and its causes. After some research, we found a sample written by a doctor[4] and adapted it in order to be 100 words long:

> If we have a dry mouth, there is not enough saliva to dissolve all the chemicals. Complete or partial loss of taste usually stems from degeneration of the taste buds due to the aging process. But the most common cause is dry mouth, which can be caused by many factors, including drugs. Damage to taste buds also arises from inflammation — stomatitis — or nasty conditions such as cancer of the mouth, or where radiotherapy has been given. The special nerves conveying the taste messages can be damaged in a head injury, tumour of the brain or after surgery on the head.

[1] Available at https://github.com/GoogleCloudPlatform/android-docs-samples
[2] Available at https://www.microsoft.com/cognitive-services/en-us/speech-api
[3] Direct download link: http://www.cambridgeenglish.org/Images/174037-first-2015-sample-papers-1.zip
[4] Full text: http://www.englishmed.com/html/reading/taste/taste.htm

8.1.3 Results

Below are the texts returned by both programs in both situations tested. When analysing them, we noticed that some errors were, in our opinion, more justifiable than others. For example, we could understand how the software could have changed a word for its homophone more easily than how it chose completely unrelated words, both in meaning and pronunciation. However, we could not find any widely used metrics to calculate the word error rate that differentiated between different types of errors. Because of that, we decided to develop our own categorisation of errors, so that we could better analyse the results of each transcription.

Hence, the texts below are exact copies of the output retrieved, with the exception of the formatting. All errors are marked in bold, italicised, or both; those distinctions correspond to the following error types:

Non-significant error - *Italics*	Misinterpretation that does not change the meaning of the sentence (examples: “river” changed to “River” or “there is” to “there’s”). These were not considered when calculating the word error rate.
Intonation error - ***Bold and italics***	Erroneous word that has the same root as the original word, may have been misspoken due to enunciation problems, and does not heavily compromise the overall understanding of the sentence. Alternatively, a homophone of the original (examples: “taste” to “tasty”, “Hale” to “hail”, or “damages” to “damaged”).
Meaningful error - **Bold**	All other errors.

Bing Speech API’s transcription of the generalist text:

> We live on the island of **Halo**. It’s about 2 kilometers? **Why did this** point **any chance** to the mainland by a narrow ***roads*** built across the mouth of the *River* **with *separated*** from the rest of the country. Most of the time **it** wouldn’t know **where** on an island because **every for months** between us. **You just lost power** and Brown. But **Windows** a high tide **in** the water rises. **Help me to Rosa about** the ***roads***. Then you know it’s definitely **not**.

Google Cloud Speech API’s transcription of the generalist text:

> we live on the island of ***hail*** it’s about ***to*** kilometres wide at its broadest ***points*** and it’s joined to the mainland by narrow ***roads*** built across the mouth of the river **with *separated*** from the rest of the country
>
> most of the time you wouldn’t know **where** on an island because the river **miles** between us and the mainland is just a **fast rate** of **Paul Russell** and **Bromley**

what is the high ***tides*** **in** the water **rides the** half a metre or so above the road then you know it's definitely an island

Bing Speech API's transcription of the medical text:

If **you** have a dry mouth. *There's* not enough **to live with yourself**. All the chemicals complete or partial **lots** of ***tasty*** **food stamps** from **this generation** of **days past** due to the aging process, but the most common ***causes*** dry mouth, which **could** be caused by many factors, including drugs. Also **rises** from **Ation Tytis** or **not ticklish and said**, **just cancelled** the mouth or **wherever therapy** has been given the special **is going to text** messages. **You tomorrow to Brian** or after surgery on the head.

Google Cloud Speech API's transcription of the medical text:

if **you** have a dry mouth there is not enough saliva to **herself** all the chemicals

little partial loss of taste usually **stands** from degeneration of the tastebuds due to the aging process

but the most common ***causes*** dry mouth which can be caused by many factors including drugs

damaged taste buds also **Rises** from inflammation stomatitis or **nice** ***condition*** **certificate over** the mouth or **a** radiotherapy has been given

special nerves **causing** the **face** messages can be damaged in **the** head injury **tomorrow after** brain or after surgery on the head

8.1.4 Analysis

To analyse the excerpts we used the word error rate (WER) metric put forward by Fiscus et al. [FAG08], which consists in "the sum of transcription errors, (word substitutions, deletions, and insertions) divided by the number of reference words, and expressed as a percentage". We also established the following rules:

1. BSA transcriptions are punctuated, whereas GCSA's are not. We chose not to evaluate the punctuation since we would have no term of comparison, and any capitalised words in the beginning of a sentence were not considered mistakes.

2. We count contractions (such as "it's" or "wouldn't") as just one word, since when spoken they behave as one due to the absence of silences or breaks when enunciating them.

3. The specific word "tastebuds" (in GCSA's transcription of the medical text) was considered correct. In the original it is spelt "taste buds", but since those are alternative spellings of the same term, we did not consider it a mistake. However, it should be noted that the same transcription also spells it as "taste buds" in another occurrence.

We found that sorting the transcription errors into the three aforementioned types helped to better measure accuracy, since we believe that the writing of a similar word to the one pronounced is a less severe error than interpreting a word that is completely unrelated in terms of meaning and sound. However, since the distinction between intonation and meaningful errors can be somewhat subjective at times, we grouped both when calculating the WER, giving separate scores only as a way of getting an additional general idea of the transcription accuracy (see table 8.1). Please note that errors of deletion (which happen when words in the recording are not registered in its transcription) are not highlighted in the transcriptions, but were still accounted for when calculating the WER. Also note that the original texts had exactly 100 words each.

Table 8.1: Word error rates of Bing Speech API and Google Cloud Speech API.

	Generalist text		Medical text	
Errors	**BSA**	**GCSA**	**BSA**	**GCSA**
Non-significant	1	0	1	0
Intonation	3	6	2	3
Meaningful	41	18	46	21
Word error rate	44%	24%	48%	24%
Ratio of intonation errors	7%	25%	4%	13%

A summary of this table can be found in Figure 8.1. Through this analysis, it became clear that GCSA had a much lower error rate. In the medical excerpt, which is arguably the most important for our evaluation, GCSA had exactly half of the error rate of BSA, while having more than three times the ratio of intonation errors. This clearly puts it in an advantage, since it will be more easily understandable by the users. It is also noteworthy to highlight GCSA's ability to comprehend medical terms like "inflammation", "stomatitis", and "taste buds", all of which were misheard by BSA.

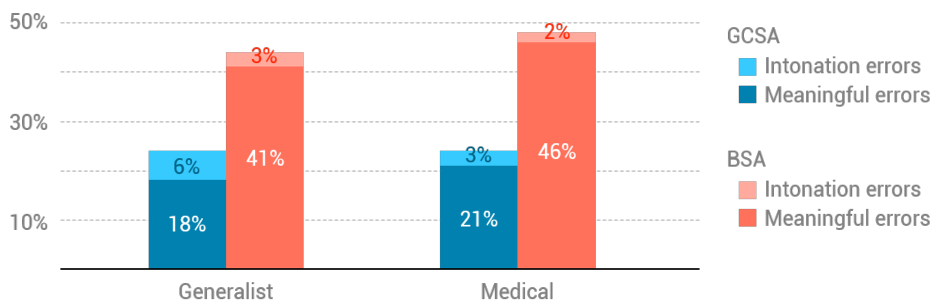


Figure 8.1: Graph comparing the word error rates of Google Cloud Speech API (blue bars) and Bing Speech API (orange bars) for both types of text (generalist and medical).

8.1.5 Discussion

The official word error rate of Google's speech technology is 4.9% and Microsoft's is 5.9% [Pro17], which are significantly lower than the 24% and 44% measured in these trials. However, it must be noted that it is not clear if those companies use the exact same speech technology for all of their speech-enabled products (which means GCSA and BSA in particular might present a different WER), and we do not know the metrics each of them is using to calculate their WER, since there are several variations (such as the ones mentioned in Section 4.1.2 between Liu et al. [LTH+11] and Devine et al. [DGC00]). Moreover, these trials were conducted with a non-native speaker, whose pronunciation may skew the results to higher WER's. Nevertheless, we feel confident that GCSA is a better solution when compared to BSA.

8.2 User Experience Assessment

To test the developed layout, five seniors volunteered to test a prototype of the app. The general consensus among UX researchers is that five users should be enough to find the vast majority of the errors in usability testing [Vir92] [TLN06]; however, since we have a very heterogeneous group of possible end users for HealthTalks, that number might not be enough to represent all users and ensure most of the errors are found [Cau01].

Since we did not have the means of implementing a comprehensible test with several diverse people, we deliberately targeted only five older people in order to gather the errors that they find with the app. Our aim was to target the group of users that we perceived as one of the less technology-oriented. In this way, if the app was intuitive and easy to use for them, the more technologically versed people should have no problems as well. However, this assumption has two flaws: the first is that we cannot know for sure if older people in general, and the five volunteers in particular, are among the less proficient with technology users. The other is that, even if they are, there may be some problems that only more proficient users may find, and those problems would remain unidentified after this test.

Even so, we did the tests as a way to ensure a coherent user experience for the older population in general. Those tests were divided into three parts: the first part consisted in measuring the testers' Technology Readiness Index (TRI); the second was an icon selection in order to choose the ones more appropriate to be used in the app; and the third was testing the human-computer interaction (HCI), done with an early prototype of the app.

8.2.1 Testers

Five senior citizens volunteered to participate in the test on the grounds of confidentiality by signing a declaration of consent (see Appendix E); for that reason, we will not disclose their names and instead refer to them as testers 1 to 5. They are characterised as such:

- **Tester 1:** Male, 66 years old, has a bachelor's degree, has an Android smartphone, and has installed new apps maybe once or twice.

- **Tester 2:** Female, 69 years old, has a bachelor's degree, has an Android smartphone, and rarely installs new apps.
- **Tester 3:** Female, 84 years old, completed the 12^{th} grade, and has an Android smartphone but has not had the time to install new apps yet since it is a recent acquisition.
- **Tester 4:** Female, 74 years old, has a 3-year degree, and does not have a smartphone.
- **Tester 5:** Female, 80 years old, has a bachelor's degree, and does not have a smartphone.

From analysing the testers' characteristics we can see that most (4/5) are female and most (4/5) have a higher education level (and the one who does not have one still completed the 12^{th} grade). They have an age range of 18 years and are split in half in terms of having a smartphone. So, even though this sample does not represent the whole elderly population (especially less educated individuals), there is still some heterogeneity among the testers.

The tests were conducted individually and presentially and the usability tests were conducted on a smartphone, as shown in Figure 8.2. For those, the smartphone's interface and the conversation between the interviewer and the testers were recorded in order to be later evaluated and compared.

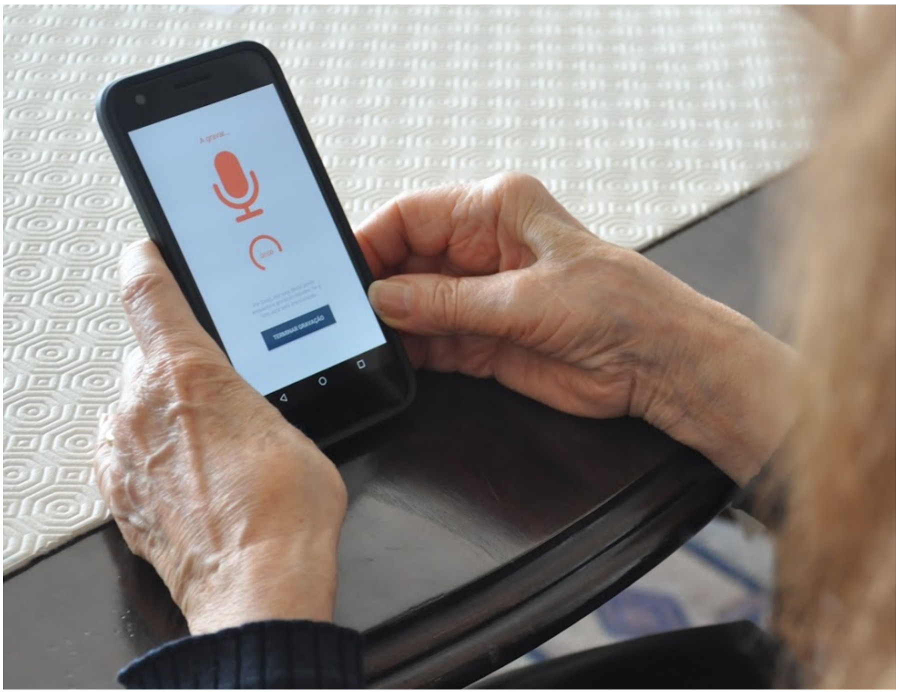


Figure 8.2: Photo of a tester during the usability test.

8.2.2 Technology Readiness Index

When planning the usability tests, we realised that the testers' openness to technology could influence the test results. This is because someone who is more receptive to trying new technologies could have an easier experience when trying our app due to their attitude towards such a solution. We wanted to compare the test results with a quantitative measure of the tester's perspective on technology in order to identify such situations and analyse the results accordingly.

To that effect, we employed the Technology Readiness Index (TRI). In 2000, Parasuraman [Par00] first proposed the concept of "technology readiness" (TR), which by his own words consists in "people's propensity to embrace and use new technologies for accomplishing goals in home life and at work". The Technology Readiness Index was obtained from a 36-item query that resulted in a classification from 1 to 5. It was created in the context of companies measuring their customers' TRI and planning their technological strategies accordingly. In 2015, Parasuraman and Colby [PC15] updated that concept to TRI 2.0, which has scales of 10 and 16 items; the latter is the version that we used in this study. Both indices are widely used in technology-oriented studies [BM10; LGG+06; LH06; DB10; WLS07].

The TRI ranks an individual in four different dimensions: optimism, a positive outlook on technology's benefits; innovativeness, the inclination to be at the forefront of technology adoption; discomfort, the feeling of being overwhelmed by technology; and insecurity, a distrust of technology and its potential harms [PC15]. While higher optimism and innovativeness lead to a higher TRI, higher levels of discomfort and insecurity lower the total score.

It works by asking the respondent to classify each of the 16 statements with a 5 item scale (from "strongly disagree" to "strongly agree"). Even though the authors give some leeway to investigators by allowing empty answers or an optional "not sure" option, we did not follow that route. But we did randomise the order of the statements as recommended, and made a small introductory text explaining the technology we were targeting: mobile devices. The statements are available in Appendix F.

Each dimension (optimism, innovativeness, discomfort and insecurity) was measured and then the final TRI score was calculated with a formula given by the authors, who kindly granted us permission to use this scale in the scope of this dissertation. In Table 8.2 we present the results of this instrument. All of the scores are measured in a scale from 1 to 5, with a higher TRI value representing a higher tendency to adopt new technology.

Table 8.2: UX tester's TRI values.

Individual	Optimism	Innovativeness	Discomfort	Insecurity	Global TRI
Tester 1	4.5	3.25	2.5	2.75	3.63
Tester 2	4.75	4.75	2.75	3	3.94
Tester 3	4	3.25	3.5	3.5	3.06
Tester 4	2.5	2	3.5	3.25	2.44
Tester 5	4.25	3	4.25	4.5	2.63

Knowing that the scale's halfway point is 3.0, then we can see a somewhat balanced distribution, with no tester having a score lower than 2.0 or higher than 4.0, as seen in Figure 8.3, and an overall average TRI of 3.14. As for specific components, Tester 4 stands out for having her optimism and innovativeness both below 3.0, unlike her peers. On the opposite side of the spectrum is Tester 2, with 4.75 in both, the highest values of all the testers. As for discomfort and insecurity, both Tester 1 and Tester 2 showed results below or equal to 3.0, while Tester 4 was the only one that presented values above 4.0. Unsurprisingly, Testers 1 and 2 had the highest overall TRI scores while Testers 4 and 5 had the lowest.

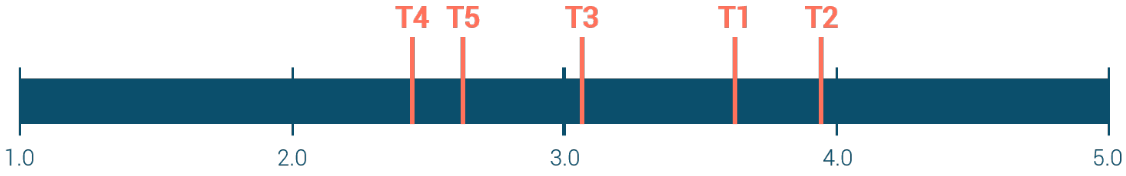


Figure 8.3: TRI 2.0 result scale with the testers' scores. Each tester is represented by the letter "T" followed by the respective number identifier.

Crossing this data with the one referred in the previous section (Section 8.2.1), we can see that Tester 4 and Tester 5, the ones with the lowest global TRI values, are the ones that do not own smartphones.

8.2.3 Iconography

One situation that happens with people who are not as used to technology is that they do not make the same icon associations as people who are more accustomed to user interfaces. Even in general, the usage of some icons may create a less clear and efficient layout; for example, the "hamburger" menu (an icon composed of three horizontal lines that is often used to show and collapse navigation menus) has been shown to reduce user engagement compared to tabs [Ros14].

An icon can be classified as a resemblance icon, which means that the symbol represented is closely related to the concept it is trying to convey (e.g.: a clock to represent time, or an envelope to symbolise e-mail); or a reference icon, which works by displaying a reference or analogy to the action associated with it (e.g.: a floppy disk to save a document, even if it was once a resemblance icon; or a flame to represent an active discussion). An icon can also be just an arbitrary convention, such as using a triangle with an exclamation point as a warning sign [Nie14].

In that regard, our app has many concepts for which there are no standard icons. The concepts of health establishment, health professional, medical appointment, and medical speciality, for example, are all very intertwined and have no clear resemblance, reference, or conventional icon that could differentiate them unequivocally. As such, it was our goal to discover what icons people most associated with those ideas, and get confirmation for the more common icons (such as the ones for date or settings) to see which were more intuitive for senior users.

Most icon tests are done in context with the rest of the interface, but that may increase uncertainty on which of those elements — the icons or the general layout — is interfering with a tester's experience. Hence, we decided to conduct isolated testing of the app's icons. Some testing techniques applicable in that context include A/B testing, which consists in the comparison between two versions of a solution [Nie05]; recognition methods, in which a tester has to guess what symbol is on the icon; and information-scent methods, during which the tester has to guess what functionality the icon stands for, regardless of the object it depicts [Har16]. However, all of those tests assume that the researcher only displays at most two possible icons to the tester, but in our case there are a lot more possibilities, especially when having to differentiate between such closely related concepts. Hence, we decided to conduct an assessment in which the testers were given several alternatives to choose from for each concept. The icons used were from Font Awesome[5] and IcoMoon[6].

The list of icons in Figure 8.4 was shown to the testers before they were allowed to see the app, since we did not want to influence their opinion by showing the icons used in the prototype. To avoid overwhelming the user, the icons were divided in different rows according to their general meaning and only one or two rows were presented at a time when asking them to choose an icon corresponding to a concept, as specified in the figure.

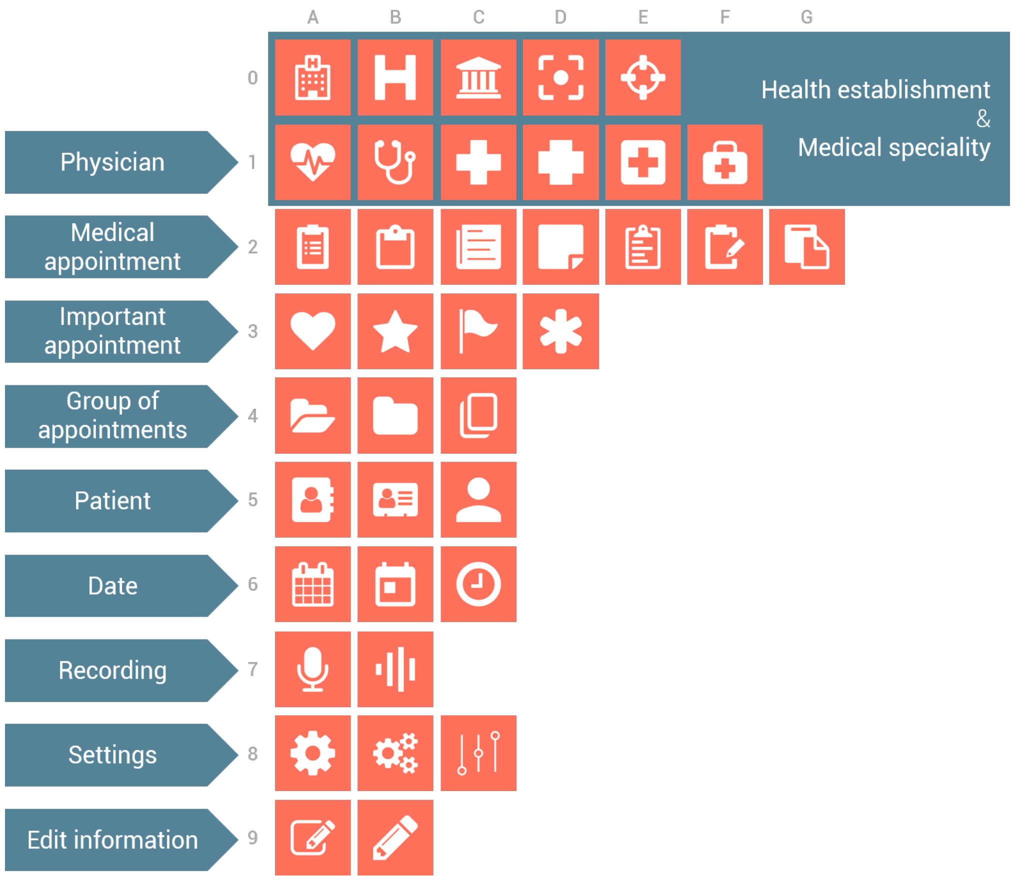


Figure 8.4: Icons shown to the testers and concepts associated with each row.

The testers were shown each row for a few seconds. However, if they did not find any icon that was suitable or took a very long time, the researcher would uncover all of the icons and ask once

[5] http://fontawesome.io/
[6] https://icomoon.io/

more for an icon to represent that concept. If the icon they chose was outside the predetermined row, the answer would be annotated but differentiated from the other answers.

The results were unanimous for the concepts of “date” (icon 6A) and “edit” (icon 9A), both of which are standard resemblance icons. Four of the testers also agreed on icons for a “group of appointments” (icon 4A), “patient” (icon 5C), and “recording” (icon 7A), which are also resemblance icons and widely used for those concepts. The icon for “physician” chosen by a simple majority of three testers was 1F, a reference icon. Other concepts’ results can be seen in Figure 8.5; the icons with more votes for “health establishment” and “settings” are resemblance icons; the icon chosen for “medical speciality” is a reference icon and the “important appointment” is an arbitrary standardised icon. The most controversial choice of icon was for the “medical appointment” concept. Two people thought the icon 2E was the most appropriate, but the other three did not feel like that was a good representation, with suggestions outside of the row including the icons 0A, 1A, 1B, and 1C.

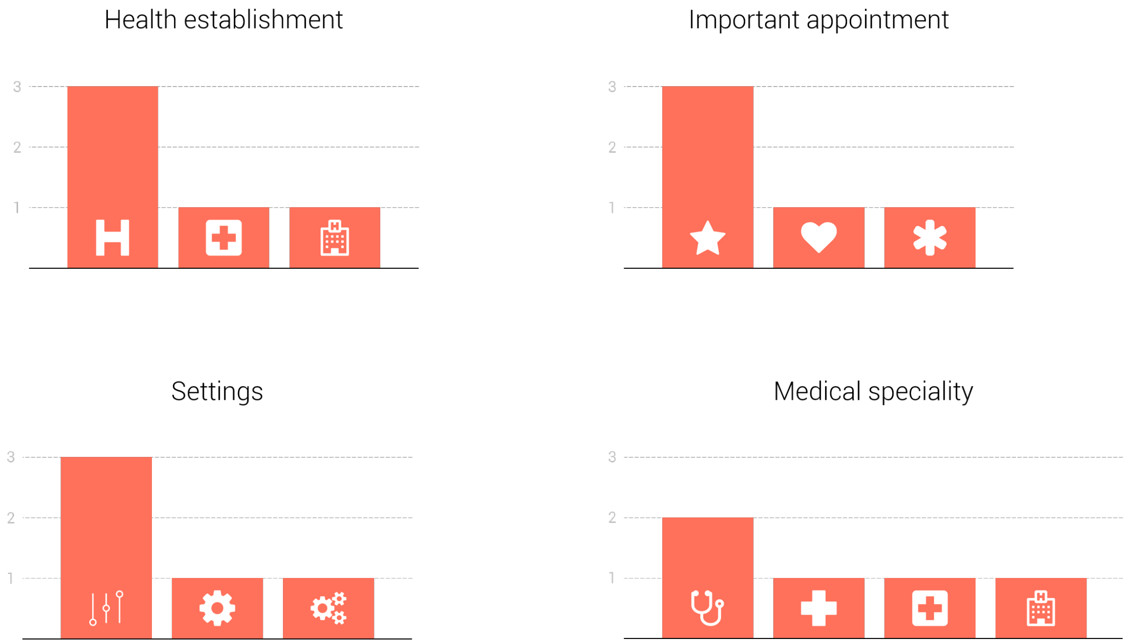


Figure 8.5: Icons chosen for the concepts of “health establishment”, “important appointment”, “settings”, and “medical speciality”.

The lack of consensus in the appointment icon, especially when juxtaposed with the general consensus in the other answers, indicates that there might have been a problem with the questioning. The icons presented to the testers in that row were mainly clipboards and papers, but the testers that did not feel like any of those were appropriate ended up choosing health-related icons such as a cross, a hospital, a stethoscope, and a beating heart. From this we can infer that they did not associate a medical appointment with a piece of paper, but rather with treatment and healing. We speculate that that divergence may have been caused by the wording of the question. In the script, there is never the mention of an app (even though the testers already knew that, ultimately, that was our project), and they were simply asked to choose an icon that represented each concept. In the general sense, it is logical that they would opt for health icons. Our choice to present them mostly clipboards and similar icons was because those are also sometimes associated with note taking by physicians during a medical appointment, and also because the concept of an appointment in the scope of HealthTalks is actually a combination of data (such as date, physician, recording, notes taken), and not so much a medical appointment in the literal sense. Because of that fact, and to avoid confusion among the other health-related concepts, we chose the icon 2E that two testers selected, a clipboard, to represent a medical appointment in the app, despite the discrepancy of the other answers.

Even though it may have restricted the options of the testers, we feel like limiting the available icons was a good decision. Some icons such as "edit" cannot be too different from the standard pencil icon so as to not alienate users with higher technology readiness and more familiarity with standardised symbols. Because of that, it was important to somewhat limit the options available to the testers. It is arguable which icons exactly should have been presented, but their limitation was, in our view, a positive aspect, even if it inevitably interfered with the testers' choices.

8.2.4 Usability test

Usability tests are an integral part of the user experience design [Shn98]. The usual procedure is to record the tester while asking them to perform a specific set of tasks. The recording may include facial expressions, cursor movements, eye patterns, dialogue, or any other data. Although the capture and analysis processes can be automatised [Ivo03], a researcher can also measure the performance and success by determining the initial and acceptance states of a test, the time that passed, the number of clicks, and many other metrics.

The moderating method used was "concurrent think aloud", which meant that the testers were incentivised to express their thoughts and reasoning while completing the tasks [US 14]. At the same time, as previously mentioned, the dialogue and smartphone screen were recorded. The disadvantage of using this methodology is that it may affect the time spent on the task or the efficiency in solving it. The test results are displayed in Table 8.3. The measured parameters were the time spent with a task, the number of clicks to reach the final state, and if the person was successful in reaching it. The tasks were asked one at a time in this specific order by the researcher: firstly, how to record an appointment; then how to check that recording after it is finished; how to create a new appointment without recording; how to edit an appointment's details; how to transcribe a recording; how to take notes on an appointment; and where more information about the app's privacy statement could be found. That same order is seen in the previously mentioned table, and was deliberately devised to introduce the tester to the functionalities in order of increasing complexity and logical usage of the app.

Table 8.3: Results of the usability test. The last column contains, for each task, both the average time spent by the successful testers in seconds, and the minimum number of clicks necessary to reach the final state.

Task	Parameter	Tester 1	Tester 2	Tester 3	Tester 4	Tester 5	Avg/Min
Record	Time (s)	99*	3	7	12	7	7.25
	Success	No*	Yes	Yes	Yes	Yes	
	Clicks	8*	1	1	1	1	1
Consult	Time (s)	4	3	7	5	6	5.00
	Success	Yes	Yes	Yes	Yes	Yes	
	Clicks	1	1	1	1	1	1
Create	Time (s)	57	38	43	40	21	39.80
	Success	Yes	Yes	Yes	Yes	Yes	
	Clicks	4	4	2	2	2	2
Edit	Time (s)	10	10	15	28	4	13.40
	Success	Yes	Yes	Yes	Yes	Yes	
	Clicks	1	1	1	1	1	1
Transcribe	Time (s)	58	9	82	7	35	33.25
	Success	No	Yes	Yes	Yes	Yes	
	Clicks	3	2	2	2	3	2
Take notes	Time (s)	3	4	4	4	5	4.00
	Success	Yes	Yes	Yes	Yes	Yes	
	Clicks	1	1	1	1	1	1
Privacy	Time (s)	6	18	36	21	38	23.80
	Success	Yes	Yes	Yes	Yes	Yes	
	Clicks	2	2	2	2	4	2

Tester 1's answer to task 1 is flagged because the tester himself, and later the record corroborated that, said that he was distracted by the conversation he was having with the researcher immediately before. It is a typical case of getting lost in an app; the user was stuck in the page where all the medical appointments were listed and there was only an option to add a new appointment without a recording. That same tester later could not transcribe a recording either, since he did not click on the appropriate tab, inside the appointment view.

Looking at the overall results, it becomes clear that tasks like creating a new appointment without a recording were generally more difficult than taking notes on an appointment, due to the difference in average time spent on each. The first one's pain point was a floating button with a "+" sign in the bottom right of the appointments listing page, which created a new appointment as the task oriented. Even though that floating button structure is typical and recommended in Material Design, most testers had trouble finding it. The most clicked button was a similar one on the top of that page, which grouped the appointments according to a specific field.

The only person who did not complete a task was Tester 1, who was the person with the second highest TRI. Tester 4, the one with the lowest TRI, never clicked more than what was strictly necessary to reach each goal and consistently took a comparable amount of time to reach it. On the time issue, it should be referred that 1) it was counted manually by watching the recordings, so it is bound to have a medium-high error rate; and 2) it often included dialogue or hesitations, which could arguably not be classified as time taken to finish the task, but which were not removed. For that reason, the time variable should not be over-analysed and a better comparison variable should be the number of clicks. On those, Tester 2 and Tester 5 have an overall excess of two clicks each, while Tester 1 has 3 (not considering the first task). Seeing as Tester 1 and Tester 2 have the highest TRI, a possible justification for a higher number of clicks may lie in their openness to experimenting and clicking fearlessly; and that same motive can justify Tester 4's minimum clicks if she felt hesitant or insecure in clicking without proper consideration.

At the end of the tests, the participants were asked if they thought HealthTalks was useful, what was its best feature, and if they would use it themselves if given the chance. The only person who said it was not a useful app was Tester 5, who justified her answer by explaining that it was not useful to her personally, since she takes notes on paper already. Everyone except Tester 2 said that the recording was HealthTalks' best feature; Tester 2 said it was the option of programming alarms. Tester 1 also mentioned the definitions of medical terms, and Tester 5 the possibility of transcribing the recordings. When asked if they would use the app, everyone said yes barring Tester 5 (who explained again that she preferred to take notes on paper) and Tester 2, who is close to her physician and hence does not need HealthTalks' functionalities.

8.2.5 Induced Changes

The most direct consequence of these tests was the icon change, which followed the testers' choices. Besides that, icons were added to the tabs in the appointment page in order to hopefully make them more understandable and visible. Moreover, the two floating buttons in the appointments listing (one to search appointments and the other one to add an appointment) were differentiated through a colour change and a more contrasting and new icon in the former, changing the ordering icon to a magnifier, which is more universal. The plus icon in the floating button to create a new appointment was also increased in size. Bigger fonts were also applied in smaller text strings throughout the app to increase their readability. The transcription button was also made more visible and highlighted in its page by changing its colour, position, and size.

It was clear from the testing that some older people have difficulty targeting specific points due to their hands shaking or their sight being imperfect. Hence, there was also a restructuring of the definitions given in a transcription. Initially, they were going to be presented in the text; a recognised concept would be highlighted in the text and, when clicked, would present its definition. However, it became clear that some elders would have trouble clicking such a small target, or they could click it accidentally when scrolling down to read the rest of the transcription. Other people still could not recognise the highlight in the medical concepts as a link, making it an ineffective source of information. Due to all of those reasons, that interface was changed so that all of the definitions were presented at the end of the page, after the transcript, in wide, individual buttons to be clearly understood and easy to select.

A few more changes should have been done if we were exclusively following the results of the tests. However, an effort was made to not overly undermine the Material Design stipulations so as to not alienate users with higher TR. Because of that, certain details that might have created problems such as the floating button remained in the interface, despite the testers' difficulty using it.

8.3 Final Speech Recognition Testing

After the implementation of the GCSA in the HealthTalks app we decided to test it with two texts in European Portuguese (one generalist and another one with medical concepts) to compare the error rate of the software between Portuguese and English. For the most part, we followed the same methodology as described in Section 8.1. However, we added one more variable in this test: the GCSA phrases. When testing this feature, we used the exact same recording file each time to avoid any variance between trials. More information is specified in Section 8.3.3.

8.3.1 Generalist Text

In order to obtain a generalist simple text in Portuguese, we looked for a sample around the CEFR B2 level, just as we did with its English counterpart in Section 8.1.1, in order to compare texts of similar difficulty. We found a text from an EPFOL (Exames de Português para Falantes de Outras Línguas, or Portuguese Exams for Speakers of Other Languages) sample paper[7] and adapted it to have exactly 100 words:

> Vou algumas vezes a Lisboa, sempre em viagens de negócios. Se não todos os meses, pelo menos de seis em seis semanas apanho um avião para Lisboa. Compreendo a língua o suficiente para não precisar de intérprete, porque uma boa parte da minha infância e adolescência foi passada no Brasil, onde os meus pais viveram alguns anos, também por razões profissionais. Conheço a cidade razoavelmente bem: os lugares onde nos deslocamos com maior frequência começam a certa altura a tornar-se familiares, pelo menos à superfície, mesmo quando, a um nível mais profundo, quase tudo neles nos faz sentir perpétuos estrangeiros.

[7] Available at https://issuu.com/lidel/docs/exames_de_portugue__s_b2

8.3.2 Medical Text

The medical text tested was an adaptation of the Wikipedia page on myasthenia gravis[8]. It was chosen due to the medical vocabulary used and because it is a thorough explanation of a single illness, thus being comparable to something a physician could say. It was slightly modified to be 100 words in length:

> A miastenia grave é uma doença caracterizada pelo funcionamento anormal da junção neuromuscular que acarreta episódios de fraqueza muscular. Na maioria dos pacientes a fraqueza é causada por anticorpos. Os sintomas variam de doente para doente, mas tipicamente podem incluir a queda de uma ou ambas as pálpebras, visão dupla, fraqueza dos músculos oculares, disfagia, dificuldade a falar, e fraqueza nos músculos da mastigação. Desconhece-se o que desencadeia o ataque do organismo contra os seus próprios receptores, mas a predisposição genética desempenha um papel essencial. Os anticorpos circulam no sangue e as mães podem passá-los ao feto através da placenta.

8.3.3 Phrases

As explained in Section 7.1.1, the GCSA allows for a small list of words or terms to be sent along with an audio file. Regardless of GCSA's own dictionary, those concepts will be given emphasis in the transcription process and will be recognised more often in the speech.

When creating the list of words to send, our concern was not to reproduce a realistic scenario, but rather measure the efficacy of that GCSA feature under ideal conditions. Consequently, we specifically sent the exact medical concepts used in the text, such as "miastenia", "disfagia", and "sangue", without considering if they had already been correctly identified without the dictionary or not. For multiple word terms, we sent all of the words both together and separately, as long as on their own they still represented a commonly used medical concept. For example, we sent both "miastenia" and "miastenia grave", but not "grave" on its own; in "predisposição genética", we sent both the whole expression and each word individually. We confined our scope to the inflections used in the text; for example, we sent "oculares" but not its singular form "ocular", since the latter is not used in the text.

The result was a collection of 22 medical expressions, most of which had already been recognised by GCSA without using this feature. However, the transcription result when sending the list was exactly the same as it had been before when no list was sent. Only four medical concepts were not being correctly identified: "miastenia grave", "predisposição genética", "sangue", and "feto"; and our list of concepts was not helpful in highlighting those terms.

After reflecting on the lack of differences between the transcriptions, we came to the conclusion that the algorithm was probably having some difficulties discerning the words from each other. For example, the text begins with the expression "a miastenia grave"; our assumption was

[8] Available at https://pt.wikipedia.org/wiki/Miastenia_grave

that the program blended the first two words together due to the lack of a pause between them; thus it did not perceive the given concept "miastenia" as a good match since it does not start with an "a" sound.

To test that hypothesis, we added five new expressions to our dictionary: "a miastenia", "a miastenia grave", "mas a predisposição genética", "circulam no sangue", and "passá-los ao feto". The results on two of those medical concepts were improved, as can be seen in Section 8.3.4 and is further discussed in Section 8.3.6.

8.3.4 Results

Following the same classifications as previously outlined in Section 8.1.3, what follows are the results from the GCSA transcription. Please keep in mind that the GCSA does not employ punctuation, hence the uninterrupted stream of words. As a reminder, the formatting of the text consists in italics for non-significant errors (not considered when calculating the WER), bold and italics for intonation errors (when the obtained word is similar in sound or root to the original one), and bold for meaningful errors (all the other errors).

Transcription of the generalist text:

> vou algumas vezes a *lisboa* sempre em viagens de negócios **em** todos os meses pelo menos de 6 em 6 semanas apanho ***o*** avião para *lisboa* compreendo a língua o suficiente para não precisar de **internet** uma boa parte da minha infância e adolescência foi passada no *brasil* **com** os meus pais viveram alguns anos também por razões profissionais ***conhece*** a cidade **de** razoavelmente bem os lugares onde nos deslocamos **comércio convencer** começam a certa altura a tornar-se familiares pelo menos ***a*** superfície mesmo quando a um nível mais profundo quase tudo **menos** nos faz sentir **vento** estrangeiros

Transcription of the medical text (without personalised phrases):

> **amigas temia** grave é uma doença caracterizada pelo funcionamento anormal da junção neuromuscular que acarreta episódios de fraqueza muscular na maioria dos pacientes a fraqueza é causada por anticorpos os sintomas variam de doente para doente mas **se** tipicamente podem incluir a queda de uma ou ambas as pálpebras visão dupla fraqueza dos músculos oculares disfagia dificuldade a falar e fraqueza nos músculos da mastigação desconhece-se o que desencadeia o ataque do organismo contra os seus próprios receptores **massapê** ***disposição*** genética desempenha um papel essencial os anticorpos ***círculo*** **mustang** e as mães podem ***passar*** **os afetos** através da placenta

Transcription of the medical text with personalised phrases:

> a miastenia grave é uma doença caracterizada pelo funcionamento anormal da junção neuromuscular que acarreta episódios de fraqueza muscular na maioria dos pacientes

a fraqueza é causada por anticorpos os sintomas variam de doente para doente mas **se** tipicamente podem incluir a queda de uma ou ambas as pálpebras visão dupla fraqueza dos músculos oculares disfagia dificuldade a falar e fraqueza nos músculos da mastigação desconhece-se o que desencadeia o ataque do organismo contra os seus próprios receptores **massapê** ***disposição*** genética desempenha um papel essencial os anticorpos ***círculo*** no sangue e as mães podem ***passar*** **os afetos** através da placenta

8.3.5 Analysis

As in Section 8.1.4, we used the WER definition defined in Fiscus et al. [FAG08], counting all the word substitutions, insertions, and deletions, though the latter are not highlighted in the transcriptions. The results can be seen in Table 8.4.

Table 8.4: Word error rates of Google Cloud Speech API with Portuguese texts.

Errors	Generalist text	Medical text	Medical text with phrases
Non-significant	3	0	0
Intonation	3	3	3
Meaningful	11	9	5
Word error rate	14%	12%	8%
Ratio of intonation errors	21%	25%	38%

The WER of the generalist and medical texts was similar, being 14% for the former and 12% for the latter. The ratio of intonation errors (which we consider less severe) is higher in the medical text, accounting for 25% of all the errors, but very close to the 21% of the generalist text. When it comes to the improvement of the medical text by using the feature of inserting custom phrases in GCSA, the differences are not very pronounced. The difference in WER is just 4%, and the text where we used phrases has a higher ratio of intonation errors since all of the corrected errors were meaningful ones.

8.3.6 Discussion

The WER we had obtained in GCSA with an English generalist text was 24%; in this generalist text we see a WER of 14%. To us, that difference seems plausible taking into account that the speaker knows Portuguese natively but not English, so a worse result was expected in the English transcription due to pronunciation errors. The same can be said of the difference between the Portuguese and English WER's for the medical speech, which were 12% and 24%, respectively.

Even though the medical text's WER is slightly lower than the one obtained in the generalist text, we do not think the difference is significant; on the contrary, we believe it gives confidence in the reliability of GCSA's performance regardless of the nature of the texts, since the WER difference between them is only 2%.

When evaluating the use of phrases in the medical text, we should note that the chosen phrases were extremely specific to the given context, including several combinations of words which are not all exclusive to the medical field. Even with such specificity, the differences between the use of phrases and their absence is, in our opinion, negligible. The WER difference is just 4%, and it was 0% before we tweaked the phrases to be much closer to the expressions used in the text, as referred in Section 8.3.3. This is probably because the GCSA already does a very good job at identifying medical concepts on its own; most of the medical or technical terms in the original medical transcription were correctly recognised. Therefore, we believe that we should not incorporate phrases in HealthTalks, but at the same time we do not think that the final solution will suffer for it. We are confident that most medical concepts will be identified, and that that result would not change significantly or at all with the input of a standard set of medical phrases in the GCSA.

As with the previous study, these samples were obtained by a researcher reading a written text, which entirely eliminates the need for filler words. These are expected to be present in a real situation with improvisation. Moreover, in a dialogue between a physician and a patient, non-lexical conversation sounds and interjections are to be expected, such as short responses or hesitation sounds, which could interfere with the WER of the transcription. Some word overlap could also exist in a medical appointment if one of the participants talks over the other. Those situations were not tested, but we intend to study how they would affect the WER in the future.

8.4 Requirement Fulfilment

In order to understand if all of the requirements initially planned for HealthTalks were fulfilled, and if so in what form, we present the same list of requirements as shown previously in Section 6.2 with the corresponding evaluation for each of them in Table 8.5 below.

Table 8.5: Requirement fulfilment.

ID	Short Description	Fulfilment
FR01	Record audio	The app records audio and saves it for later play back.
FR02	Transcribe audio	The app uses GCSA to transcribe the recorded audio.
FR03	Meanings of medical concepts	The app uses MWA to obtain a predetermined list of concepts with the corresponding definitions, and finds them in the transcriptions.
FR04	Search	The user can search for an appointment by writing a string and choosing which fields to find that string on.
FR05	Take notes	The user can take notes on each individual appointment.

SR01	Unambiguous recording	When recording, the device stays unlocked and the screen displays a microphone icon; when the user leaves that window or locks the device, the recording stops.
SR02	Privacy	The app has a privacy statement that explains how the user's data is handled and displays the contact for the Portuguese Data Security Protection National Centre; the app uses gRPC when communicating with GCSA to increase data security online.
SR03	User responsibilities	The user is made aware of the ethical consequences of being able to record another person in a text inside the app.
NR01	Simple & intuitive	The app was built with the UX in mind, and tests were done to ascertain its usability by elderly people.
NR02	Two clicks to record	Currently the recording window can be accessed by two clicks (logo to open the app, and button to record) or a long press and a click (app shortcut).
NR03	One click to stop recording	The recording stops with one click or whenever the user leaves the recording window.
NR04	Explain functionalities	The app's functionalities are explained in a text inside the app.
NR05	Compatibility	The app is compatible with Android 4.0.3, released in 2011; this means it is compatible with 99,2% of all Android devices currently in use [Goo17a].
NR06	STT through API	GCSA is used by making requests to its API.
NR07	Truthful and recent information	The concept database retrieved from MWA can be adapted to only accept pages recently modified; moreover, the database is refreshed on the device whenever HealthTalks receives an update, so as long as the app is supported, so is the database. On the veracity of the definitions, Wikipedia is considered mostly accurate and is updated with some regularity (see Section 7.1.2 for more information).
NR08	WER of 41.5% or lower	The tests of GCSA in Portuguese resulted in WER's of 12% and 14%, although the app was not tested in a real scenario.

NR09	Most functionalities work offline	The only online functionality is the transcription of an appointment. That can be done at any time by the user, and a confirmation prompt will appear if the device is not connected to a Wi-Fi network.
NR10	Free to use	The app is not yet available to the public, but there are currently no plans to monetise it.

Chapter 9

Conclusion and Future Work

In order to provide good healthcare to everyone, a lot of issues have to be addressed. We researched at length many of the topics relevant to that goal, including the state of health literacy, the doctor-patient relationship, the communication skills of physicians, how patients retrieve health information, and how some technologies may ease the burden of self-care on their users. On top of that, we researched many of those aspects within our community, which presented new and compelling data on how that population thinks and behaves.

The combination of all that information allowed us to develop and improve a new tool to help health consumers. HealthTalks is a user-driven app that provides several functionalities that its users can employ in their daily lives in order to improve their personal health information management skills, the understanding they have of their doctors, and, in the long term, possibly their health literacy.

Even though the current app is useful and up to our expectations, we intend to improve it even further in the future, continuing our efforts of improving the health condition of patients.

9.1 Objective fulfilment

In Section 8.4 we analysed each of the requirements we initially proposed for HealthTalks' development and their final state of completion; from there we can infer that all of its objectives were achieved successfully. Furthermore, we had the opportunity of testing the app with senior citizens in order to make adjustments suitable to that target population, and focused on achieving an optimal user experience since its inception. We believe that those goals were well achieved due to the generally positive results we had in the usability tests.

The great care we had in developing the app is also visible in the process we created to pick a software recognition software to use. Before applying the speech-to-text tests we were not expecting such a significant difference in word error rates; however, the fact that we did test them before allowed us to make a much more informed decision. We trust that our final choice was the best we could have made, which elevates the quality of HealthTalks.

The questionnaire was planned very early during this dissertation's development, but evolved to have a bigger scope than initially anticipated. With more than 1200 answers, we find it to be a conclusive and comprehensive study on Porto residents' attitudes and habits towards healthcare. Its results were beneficial for our solution, since they allowed us to gain a better perspective on its most wanted features and the biggest needs of the population. Our hope is that it can also be a source of information for other health-related projects in our region.

Overall, even though this project was a substantial endeavour, it was completed successfully and thoroughly. We are satisfied with its results, and we hope the app can be put to more tests and challenges in the future in order to verify its applicability in the real world.

9.2 Future work

Now that we have an exhaustive resource on how Porto's patients feel and behave, we want to change our focus to the physicians. From our limited experience with them so far, ethical concerns seem to be paramount among them. However, we want to methodically study and analyse their perspective so that we can compare it to the patients' point of view and draw conclusions from the resulting data. HealthTalks is tailored to the patients, but its use will ultimately be decided by their doctors.

Another way of progressing forward is continuing the implementation of more features in the app. Many ideas were suggested in Chapter 5 by respondents, so there is no lack of possibilities.

Finally, we wish to expand our research by evaluating other aspects of the app, such as its usability in a real-world scenario and how other segments of the target population interact with it.

References

[Aro03] Neeraj K. Arora. Interacting with cancer patients: the significance of physicians' communication behavior. *Social Science & Medicine*, 57(5):791–806, September 2003. ISSN: 02779536. DOI: 10.1016/S0277-9536(02)00449-5.

[AS07] I. Elaine Allen and Christopher A. Seaman. Likert Scales and Data Analyses. *Quality Progress*, 40(7):64–65, 2007.

[Bak06] David W. Baker. The meaning and the measure of health literacy. *Journal of General Internal Medicine*, 21(8):878–883, August 2006. ISSN: 08848734. DOI: 10.1111/j.1525-1497.2006.00540.x.

[BBC05] BBC News. Wikipedia survives research test. *BBC News*, 2005. URL: http://news.bbc.co.uk/2/hi/technology/4530930.stm.

[BBD13] Jessica De Bord, Wylie Burke, and Denise M. Dudzinski. Confidentiality, 2013. URL: https://depts.washington.edu/bioethx/topics/confiden.html (visited on 02/03/2017).

[BM10] S. Bennett and K. Maton. Beyond the 'digital natives' debate: Towards a more nuanced understanding of students' technology experiences. *Journal of Computer Assisted Learning*, 26(5):321–331, 2010. ISSN: 02664909. DOI: 10.1111/j.1365-2729.2010.00360.x.

[Bro16] Mark Brohan. Why health apps don't always deliver as promised, December 2016. URL: https://www.internethealthmanagement.com/2016/12/06/health-apps-dont-always-deliver-promised/.

[Cam15] Cambridge English. The Cambridge English Scale explained. Technical report, Cambridge English Language Assessment, Cambridge, 2015. URL: http://www.cambridgeenglish.org/images/177867-the-methodology-behind-the-cambridge-english-scale.pdf.

[Car06] Richard H. Carmona. Health literacy: A national priority. *Journal of General Internal Medicine*, 21(8):803–803, August 2006. ISSN: 0884-8734. DOI: 10.1111/j.1525-1497.2006.00569.x. URL: http://link.springer.com/10.1111/j.1525-1497.2006.00569.x.

[Cau01] David A. Caulton. Relaxing the homogeneity assumption in usability testing. *Behaviour & Information Technology*, 20(1):1–7, 2001. ISSN: 0144-929X. DOI: 10.1080/01449290010020648. URL: http://www.tandfonline.com/doi/abs/10.1080/01449290010020648.

[CH01] R. J. W. Cline and K. M. Haynes. Consumer health information seeking on the Internet: the state of the art. *Health education research*, 16(6):671–92, December 2001. ISSN: 0268-1153. DOI: 10.1093/HER/16.6.671. URL: http://www.ncbi.nlm.nih.gov/pubmed/11780707.

[Coh04] Mike Cohn. *User stories applied : for agile software development*. Addison-Wesley, 2004, page 268. ISBN: 0321205685.

[Com17] Comissão Nacional de Protecção de Dados. 10 Medidas para Preparar a Aplicação do Regulamento Europeu de Proteção de Dados. Technical report, 2017.

[Cor17] Evernote Corporation. Evernote, 2017. URL: https://evernote.com/intl/pt/.

[CRC07] Alan Cooper, Robert Reimann, and David Cronin. *About Face 3: The Essentials of Interaction Design*. Wiley Publishing, Inc, Indianapolis, Indiana, 3rd edition, 2007. ISBN: 978-0-470-08411-3.

[DB10] Catherine Demangeot and Amanda J. Broderick. Consumer Perceptions of Online Shopping Environments. *Psychology & Marketing*, 30(6):461–469, 2010. ISSN: 01482963. DOI: 10.1002/mar. arXiv: arXiv:1011.1669v3. URL: http://eprints.lancs.ac.uk/23431/.

[DGC00] Eric G. Devine, Stephan A. Gaehde, and Arthur C. Curtis. Comparative evaluation of three continuous speech recognition software packages in the generation of medical reports. *Journal of the American Medical Informatics Association: JAMIA*, 7(5):462–8, 2000. ISSN: 1067-5027. URL: http://www.ncbi.nlm.nih.gov/pubmed/10984465%20http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC79041.

[EÁM15] Rita Espanha, Patrícia Ávila, and Rita Veloso Mendes. Literacia em Saúde em Portugal - Relatório Síntese. Technical report, Fundação Calouste Gulbenkian & Centro de Investigação e Estudos de Sociologia/IUL, 2015. URL: https://gulbenkian.pt/publication/literacia-em-saude-em-portugal/.

[Eur16] European Parliament. Directive on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, 2016.

[Eva14] Kevin W Eva. Trending in 2014: Hippocrates. *Medical Education*, 48(1):1–3, January 2014. ISSN: 03080110. DOI: 10.1111/medu.12392. URL: http://doi.wiley.com/10.1111/medu.12392.

[FAG08] Jonathan G. Fiscus, Jerome Ajot, and John S. Garofolo. The Rich Transcription 2007 Meeting Recognition Evaluation. In Rainer Stiefelhagen, Rachel Bowers, and Jonathan Fiscus, editors, *Multimodal Technologies for Perception of Humans*, part The Rich T, pages 373–389. Springer-Verlag, 2008. ISBN: 978-3-540-68584-5. DOI: 10.1007/978-3-540-68585-2_36. URL: http://dx.doi.org/10.1007/978-3-540-68585-2%7B%5C_%7D36.

[FAR+13] Flávio Ferreira, Nuno Almeida, Ana Filipa Rosa, André Oliveira, António Teixeira, and José Casimiro Pereira. Multimodal and adaptable medication assistant for the elderly: A prototype for interaction and usability in smartphones. In *8th Iberian Conference on Information Systems and Technologies (CISTI 2013)*, volume 8, pages 1–6, 2013. ISBN: 978-989-96247-9-5.

[FF03] Susannah Fox and Deborah Fallows. Internet Health Resources. In *TPRC*. Pew Internet & American Life Project, 2003. URL: https://ssrn.com/abstract=2054071.

[Fox11] Susannah Fox. The Social Life of Health Information. Technical report, Pew Research Center's Internet & American Life Project, Washington, D.C., 2011, page 45.

[GJ14] Alex Graves and Navdeep Jaitly. Towards End-to-End Speech Recognition with Recurrent Neural Networks. In *Proceedings of the 31st International Conference on Machine Learning*, Beijing, China, 2014.

[Glo15] Global Network of WHO Collaborating Centres for Bioethics. *Global Health Ethics - Key Issues*. 2015. ISBN: 978 92 4 069403 3.

[GMR+98] Cheryl L. Grady, Anthony R. McIntosh, M. Natasha Rajah, and Fergus I. M. Craik. Neural correlates of the episodic encoding of pictures and words. In *Proceedings of the National Academy of Sciences of the United States of America*, pages 2703–2708, 1998.

[Goo17a] Google. Dashboards, 2017. URL: https://developer.android.com/about/dashboards/index.html (visited on 06/20/2017).

[Goo17b] Google. Developer Guide - Protocol Buffers, 2017. URL: https://developers.google.com/protocol-buffers/docs/overview (visited on 06/20/2017).

[Gri10] Pamela Grimm. Social Desirability Bias. In *Wiley International Encyclopedia of Marketing*. John Wiley & Sons, Ltd, Chichester, UK, December 2010. DOI: 10.1002/9781444316568.wiem02057. URL: http://doi.wiley.com/10.1002/9781444316568.wiem02057.

[Gur12] Burkay Gur. *Improving Speech Recognition Accuracy for Clinical Conversations*. M.Sc. Thesis, Massachusetts Institute of Technology, 2012.

[Har16] Aurora Harley. Usability Testing of Icons, 2016. URL: https://www.nngroup.com/articles/icon-testing/ (visited on 05/20/2017).

[HBS14] Thomas J. Hwang, Florence T. Bourgeois, and John D. Seeger. Drug Safety in the Digital Age. *New England Journal of Medicine*, 370(26):2460–2462, June 2014. DOI: 10.1056/NEJMp1401767. URL: http://www.nejm.org/doi/10.1056/NEJMp1401767.

[Hea17] Health Care Labs B.V. Patient Journey App, 2017. URL: https://patientjourneyapp.com/.

[HES00] Hynek Hermansky, Daniel P. W. Ellis, and Sangita Sharma. Tandem Connectionist Feature Extraction for Conventional HMM Systems. In *2000 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings*. IEEE, 2000. DOI: 10.1109/ICASSP.2000.862024.

[HL10] Jennifer Fong Ha and Nancy Longnecker. Doctor-patient communication: a review. *The Ochsner journal*, 10(1):38–43, 2010. ISSN: 1524-5012. URL: http://www.ncbi.nlm.nih.gov/pubmed/21603354%20http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3096184.

[HLS15] Inkyoung Hur, Ronald Lee, and J.J. Schmidt. How Healthcare Technology Shapes Health Literacy? A Systematic Review. In *AMCIS 2015 Proceedings*, 2015. URL: http://aisel.aisnet.org/amcis2015/HealthIS/GeneralPresentations/3.

[Hol08] Lucy Holman Rector. Comparison of Wikipedia and other encyclopedias for accuracy, breadth, and depth in historical articles. *Reference Services Review*, 36(1):7–22, February 2008. DOI: 10.1108/00907320810851998. URL: http://www.emeraldinsight.com/doi/10.1108/00907320810851998.

[HUR+15] Monique Heijmans, Ellen Uiters, Tamsin Rose, Jolien Hofstede, Walter Devillé, Iris van der Heide, Hendriek Boshuisen, and Jany Rademakers. Study on sound evidence for a better understanding of health literacy in the European Union. Technical report, Consumers, Health and Food Executive Agency (Chafea), Luxembourg, 2015. DOI: 10.2818/150402.

[Hus16] Husobee. REST v. gRPC, 2016. URL: https://husobee.github.io/golang/rest/grpc/2016/05/28/golang-rest-v-grpc.html.

[HZ10] Hongbing Hu and Stephen A. Zahorian. Dimensionality Reduction Methods for HMM Phonetic Recognition. In *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2010. DOI: 10.1109/ICASSP.2010.5495130.

[Ins12] Instituto Nacional de Estatística. Censos 2011. Technical report, Instituto Nacional de Estatística, 2012. URL: https://www.ine.pt/xportal/xmain?xpid=INE%7B%5C&%7Dxpgid=ine%7B%5C_%7Ddestaques%7B%5C&%7DDESTAQUESdest%7B%5C_%7Dboui=107624784%7B%5C&%7DDESTAQUEStema=55466%7B%5C&%7DDESTAQUESmodo=2.

[Ivo03] Melody Y. Ivory. Usability Testing Methods. In *Automated Web Site Evaluation*, part II, pages 23–37. Springer Netherlands, 2003. ISBN: 978-94-017-0375-8. DOI: 10.1007/978-94-017-0375-8_3.

[JBO10] Joanne E. Jordan, Rachelle Buchbinder, and Richard H. Osborne. Conceptualising health literacy from the patient perspective. *Patient education and counseling*, 79(1):36–42, April 2010. ISSN: 1873-5134. DOI: 10.1016/j.pec.2009.10.001. URL: http://www.ncbi.nlm.nih.gov/pubmed/19896320.

[JLL+14] Maree Johnson, Samuel Lapkin, Vanessa Long, Paula Sanchez, Hanna Suominen, Jim Basilakis, and Linda Dawson. A systematic review of speech recognition technology in health care. *BMC Medical Informatics and Decision Making*, 14(1):94, December 2014. DOI: 10.1186/1472-6947-14-94.

[Jon07] William Jones. Personal Information Management. *Annual Review of Information Science and Technology*, 41(1):453–504, 2007. ISSN: 00664200. DOI: 10.1002/aris.2007.1440410117. URL: http://doi.wiley.com/10.1002/aris.2007.1440410117.

[JR04] B H Juang and Lawrence R Rabiner. Automatic Speech Recognition – A Brief History of the Technology Development, 2004.

[KCS09] Aniket Kittur, Ed H Chi, and Bongwon Suh. What's in Wikipedia ? Mapping Topics and Conflict Using Socially Annotated Category Structure. *Chi*:1509–1512, 2009. DOI: 10.1145/1518701.1518930.

[KGJ+06] Mark Kutner, Elizabeth Greenberg, Ying Jin, and Christine Paulsen. The Health Literacy of America's Adults: Results From the 2003 National Assessment of Adult Literacy. Technical report, National Center for Education Statistics, Washington, DC, 2006.

[KMG+14] Jona Kräenbring, Tika Monzon Penza, Joanna Gutmann, Susanne Muehlich, Oliver Zolk, Leszek Wojnowski, Renke Maas, Stefan Engelhardt, and Antonio Sarikas. Accuracy and Completeness of Drug Information in Wikipedia: A Comparison with Standard Textbooks of Pharmacology. *PLoS ONE*, 9(9):e106930, September 2014.

Christian Lovis, editor. DOI: 10.1371/journal.pone.0106930. URL: http://dx.plos.org/10.1371/journal.pone.0106930.

[Lau16] Virginia Lau. As users move to mobile, Google says they made 200 million queries for info about cancer drugs. *Medical Marketing and Media*, 2016. URL: http://www.mmm-online.com/technology/as-users-move-to-mobile-google-says-they-made-200-million-queries-for-info-about-cancer-drugs/article/576571/.

[LB14] MaryJo Ludwig and Wylie Burke. Physician-Patient Relationship, 2014. (Visited on 02/03/2017).

[LGF+15] Alexandre La Luna, Bruno da Silva Guilhermino, Karine Laís Fragoso, Gabrielli Lindaura Soares Medeiros, Vitória Clara de Lima da Rocha, and Izabella Bianca Almeida da Silva. O aplicativo MyFitnessPal como recurso promotor da aprendizagem de biologia. *Revista Interdisciplinar de Tecnologias e Educação*, 1:259–262, 2015.

[LGG+06] Veronica Liljander, Filippa Gillberg, Johanna Gummerus, and Allard van Riel. Technology readiness and the evaluation and adoption of self-service technologies. *Journal of Retailing and Consumer Services*, 13(3):177–191, 2006. ISSN: 09696989. DOI: 10.1016/j.jretconser.2005.08.004.

[LH06] Jiun-Sheng Chris Lin and Pei-ling Hsieh. The role of technology readiness in customers' perception and adoption of self-service technologies. *International Journal of Service Industry Management*, 17(5):497–517, 2006. ISSN: 0956-4233. DOI: 10.1108/09564230610689795. URL: http://www.emeraldinsight.com/doi/10.1108/09564230610689795.

[Lib08] Ronald T. Libby. *The Criminalization of Medicine: America's War on Doctors*. Praeger Publishers, 2008. ISBN: 978-0313345463.

[Lob13] Joana Polónia Lobo. *Personal Medication Advisor*. M.Sc. Thesis, Universidade do Porto, 2013. URL: http://hdl.handle.net/10216/67687.

[LTH+11] Feifan Liu, Gokhan Tur, Dilek Hakkani-Tür, and Hong Yu. Towards spoken clinical-question answering: evaluating and adapting automatic speech-recognition systems for spoken clinical questions. *Journal of the American Medical Informatics Association: JAMIA*, 18(5):625–30, 2011. ISSN: 1527-974X. DOI: 10.1136/amiajnl-2010-000071. URL: http://www.ncbi.nlm.nih.gov/pubmed/21705457%20http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3168303.

[MHO+15] Joyce A Martin, Brady E Hamilton, Michelle J K S Osterman, Sally C Curtin, and T J Mathews. National Vital Statistics Reports Volumne 64, Number 1, January 15, 2015. *National Vital Statistics Reports*, 64(1), 2015. URL: https://www.cdc.gov/nchs/data/nvsr/nvsr64/nvsr64%7B%5C_%7D01.pdf.

[MLP+03] Elizabeth Murray, Bernard Lo, Lance Pollack, Karen Donelan, Joe Catania, Martha White, Kinga Zapert, and Rachel Turner. The Impact of Health Information on the Internet on the Physician-Patient Relationship. *Archives of Internal Medicine*, 163(14):1727, July 2003. ISSN: 0003-9926. DOI: 10.1001/archinte.163.14.1727. URL: http://archinte.jamanetwork.com/article.aspx?doi=10.1001/archinte.163.14.1727.

[MyF17] Inc. MyFitnessPal. MyFitnessPal, 2017. URL: https://www.myfitnesspal.com/pt.

[Nal16] Maurizio Naldi. Concentration in the mobile operating systems market, 2016.

[Nie05] Jakob Nielsen. Putting A/B Testing in Its Place, 2005. URL: https://www.nngroup.com/articles/putting-ab-testing-in-its-place/ (visited on 05/20/2017).

[Nie14] Jakob Nielsen. Icon Classification: Resemblance, Reference, and Arbitrary Icons, 2014. URL: https://www.nngroup.com/articles/classifying-icons/ (visited on 06/04/2017).

[Nie94] Jakob Nielsen. *Usability Engineering*. Elsevier, editor. Morgan Kaufmann, San Francisco, CA, USA, 1st edition, 1994. ISBN: 0125184050.

[OBr11] Terrence O'Brien. Roboto font and the new design philosophy of Android 4.0, Ice Cream Sandwich, 2011. URL: https://www.engadget.com/2011/10/18/roboto-and-the-new-design-philosophy-of-android-4-0-ice-cream-s/ (visited on 05/23/2017).

[Off08] Office of Disease Prevention and Health Promotion. America's Health Literacy: Why We Need Accessible Health Information, 2008. URL: https://health.gov/communication/literacy/issuebrief/ (visited on 01/09/2017).

[Off16] Office of Disease Prevention and Health Promotion. Health Literacy Basics, 2016. URL: https://health.gov/communication/literacy/quickguide/factsbasic.htm (visited on 11/28/2016).

[Ogi16] Hamish Ogilvy. gRPC and the displacement of REST based APIs, 2016. URL: https://www.sajari.com/blog/grpc-and-displacement-of-rest-apis (visited on 06/20/2017).

[Osb13] Helen Osborne. *Health Literacy From A to Z: Practical Ways to Communicate Your Health Message*. Jones & Bartlett Learning, Burlington, MA, second edition, 2013. ISBN: 0-7637-4550-2.

[Par00] A. Parasuraman. Technology Readiness Index (Tri). *Journal of Service Research*, 2(4):307–320, 2000. ISSN: 1094-6705. DOI: 10.1177/109467050024001. URL: http://journals.sagepub.com/doi/10.1177/109467050024001.

[PC15] A. Parasuraman and Charles L. Colby. An Updated and Streamlined Technology Readiness Index. *Journal of Service Research*, 18(1):59–74, 2015. ISSN: 1094-6705. DOI: 10.1177/1094670514539730. URL: http://journals.sagepub.com/doi/10.1177/1094670514539730.

[PE15] Patient information Publications and Egton Medical Information Systems Ltd. Patient Access, 2015. URL: http://patient.info/patient-access (visited on 02/02/2017).

[PE16] Patient information Publications and Egton Medical Information Systems Ltd. Patient, 2016. URL: http://patient.info/ (visited on 02/02/2017).

[PKS04] Ronaldo Parente, Ned Kock, and John Sonsini. An analysis of the implementation and impact of speech-recognition technology in the healthcare sector. *Perspectives in health information management*, 1:5, June 2004. ISSN: 1559-4122. URL: http://www.ncbi.nlm.nih.gov/pubmed/18066385%20http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC2047322.

[PR00] Steven D. Pearson and Lisa H. Raeke. Patients' trust in physicians: Many theories, few measures, and little data. *Journal of General Internal Medicine*, 15(7):509–513, July 2000. ISSN: 0884-8734. DOI: 10.1046/j.1525-1497.2000.11002.x. URL: http://link.springer.com/10.1046/j.1525-1497.2000.11002.x.

[Pro17] Emil Protalinski. Google's speech recognition technology now has a 4.9% word error rate, 2017. URL: https://venturebeat.com/2017/05/17/googles-speech-recognition-technology-now-has-a-4-9-word-error-rate/ (visited on 05/18/2017).

[PUC+06] Wanda Pratt, Kenton Unruh, Andrea Civan, and Meredith Skeels. Personal Health Information Management. *Communications of the ACM*, 49(1), 2006.

[Red17] Rede Lusófona para a Promoção da Literacia para a Saúde. Literacia para a Saúde, 2017. URL: http://literacia-saude.info/ (visited on 01/28/2017).

[RJ86] L. R. Rabiner and B. H. Juang. An Introduction to Hidden Markov Models. *IEEE ASSP Magazine*:4–16, 1986.

[RJ93] Lawrence R. Rabiner and B. H. (Biing-Hwang) Juang. *Fundamentals of speech recognition*. PTR Prentice Hall, 1993. ISBN: 0130151572.

[RL] J. Mark Ruscin and Sunny A. Linnebur. Aging and Drugs. URL: http://www.msdmanuals.com/home/older-people's-health-issues/aging-and-drugs/aging-and-drugs (visited on 04/23/2017).

[RM13] B Raghavendhar Reddy and E Mahender. Speech to Text Conversion using Android Platform. *International Journal of Engineering Research and Applications*, 3(1), 2013.

[RMM+12] N. J. Reavley, A. J. Mackinnon, A. J. Morgan, M. Alvarez-Jimenez, S. E. Hetrick, E. Killackey, B. Nelson, R. Purcell, M. B. H. Yap, and A. F. Jorm. Quality of information sources about mental disorders: a comparison of Wikipedia with centrally controlled web and printed sources. *Psychological Medicine*, 42(08):1753–1762, August 2012. DOI: 10.1017/S003329171100287X. URL: http://www.journals.cambridge.org/abstract%7B%5C_%7DS003329171100287X.

[Rod13] João António Santos Gomes Rodrigues. *Speech-to-speech translation to support medical interviews*. M.Sc. Thesis, Universidade de Lisboa, 2013. URL: http://hdl.handle.net/10451/10731.

[Ros14] Anthony Rose. UX designers: Side drawer navigation could be costing you half your user engagement. *The Next Web*, 2014. URL: https://thenextweb.com/dd/2014/04/08/ux-designers-side-drawer-navigation-costing-half-user-engagement/%7B%5C#%7D.tnw%7B%5C_%7DXvnlggZS.

[RPE+98] Thomas N. Robinson, Kevin Patrick, Thomas R. Eng, and David Gustafson. An Evidence-Based Approach to Interactive Health Communication - A Challenge to Medicine in the Information Age. *JAMA*, 280(14):1264, October 1998. ISSN: 0098-7484. DOI: 10.1001/jama.280.14.1264. URL: http://jama.jamanetwork.com/article.aspx?doi=10.1001/jama.280.14.1264.

[SB02] Ken Schwaber and Mike Beedle. *Agile Software Development with Scrum*. Prentice Hall, Upper Saddle River, 1st edition, 2002.

[SDN+16] K. Singh, K. Drouin, L. P. Newmark, J. Lee, A. Faxvaag, R. Rozenblum, E. A. Pabo, A. Landman, E. Klinger, and D. W. Bates. Many Mobile Health Apps Target High-Need, High-Cost Populations, But Gaps Remain. *Health Affairs*, 35(12):2310–2318, December 2016. ISSN: 0278-2715. DOI: 10.1377/hlthaff.2016.0578. URL: http://content.healthaffairs.org/cgi/doi/10.1377/hlthaff.2016.0578.

[Shn98] Ben Shneiderman. *Designing the User Interface – Strategies for Effective Human-Computer-Interaction*. Addison-Wesley Longman, 3rd edition, 1998.

[Sou16] Hugo Miguel Ribeiro de Sousa. *HealthTranslator: automatic annotation of Web documents in order to assist health consumer's searches*. M.Sc. Thesis, Universidade do Porto, 2016.

[SVF+12] Kristine Sørensen, Stephan Van den Broucke, James Fullam, Gerardine Doyle, Jürgen Pelikan, Zofia Slonska, and Helmut Brand. Health literacy and public health: A systematic review and integration of definitions and models. *BMC Public Health*, 12(1):80, 2012. ISSN: 1471-2458. DOI: 10.1186/1471-2458-12-80. URL: http://bmcpublichealth.biomedcentral.com/articles/10.1186/1471-2458-12-80.

[SZH+15] Hanna Suominen, Liyuan Zhou, Leif Hanlen, and Gabriela Ferraro. Benchmarking clinical speech recognition and information extraction: new data, methods, and evaluations. *JMIR medical informatics*, 3(2):e19, April 2015. DOI: 10.2196/medinform.4321. URL: http://www.ncbi.nlm.nih.gov/pubmed/25917752%20http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4427705.

[TJB06] Jaime Teevan, William Jones, and Benjamin B. Bederson. Personal Information Management. *Communications of the ACM*, 49(1):40–43, 2006.

[TLN06] C W Turner, James R Lewis, and Jakob Nielsen. Determining usability test sample size. *International Encyclopedia of Ergonomics and Human Factors 2nd edition*, 3(April):3084–3088, 2006.

[US 14] U.S. Department of Health and Human Services. Running a Usability Test, May 2014. URL: https://www.usability.gov/how-to-and-tools/methods/running-usability-tests.html (visited on 05/21/2017).

[Vir92] R A Virizi. Refining the test phase of usability evaluation: how many subjects is enough? *Human Factors*, 34(4):457–468, 1992. ISSN: 0018-7208. DOI: 10.1177/001872089203400407.

[WC13] Emily Walsh and Ilseung Cho. Using Evernote as an Electronic Lab Notebook in a Translational Science Laboratory. *Journal of Laboratory Automation*, 18(3):229–234, June 2013. ISSN: 2211-0682. DOI: 10.1177/2211068212471834. URL: http://journals.sagepub.com/doi/10.1177/2211068212471834.

[WLS07] Rita Walczuch, Jos Lemmink, and Sandra Streukens. The effect of service employees' technology readiness on technology acceptance. *Information and Management*, 44(2):206–215, 2007. ISSN: 03787206. DOI: 10.1016/j.im.2006.12.005.

[Wor10] World Health Organization. Process of translation and adaptation of instruments, 2010. URL: http://www.who.int/substance%7B%5C_%7Dabuse/research%7B%5C_%7Dtools/translation/en/ (visited on 06/19/2017).

[Zha02] Haixia Zhao. Fitts' Law: Modeling Movement Time in HCI, 2002. URL: https://pdfs.semanticscholar.org/1337/1b377afc61d4dc918cd69edf66e78a834067.pdf.

REFERENCES

Appendix A

Questionnaire Prototype

The first version of the questionnaire had three pages, and they can be seen from Figure A.1 to A.3.

HEALTH
TALKS

Este questionário foi desenvolvido no âmbito de uma dissertação de mestrado na Faculdade de Engenharia da Universidade do Porto e pretende aferir a eficácia da comunicação médico-paciente e dos métodos de gestão pessoal de informação de saúde. O inquérito demora aproximadamente 5 minutos a responder.

A sua identidade permanecerá anónima.

Se quiser esclarecer alguma dúvida pode contactar o mestrando João Monteiro através do seguinte endereço de correio eletrónico: ei11055@fe.up.pt

Quando terminar, por favor coloque o questionário na caixa para o efeito que se encontra na receção. Obrigado pela sua colaboração!

I. Dados demográficos e outros dados

Género

- O Masculino
- O Feminino
- O Outro: __________

Idade

- O 18-25 anos
- O 26-35 anos
- O 36-49 anos
- O 50-64 anos
- O 65+ anos

Habilitações (por favor, selecione o grau de ensino mais elevado que tenha completado)

- O Não completei a 4.ª classe
- O 4.ª classe
- O 9.º ano
- O 12.º ano
- O Licenciatura ou bacharelato
- O Mestrado
- O Doutoramento

Possui um smartphone ou tablet?

- O Sim →
- O Não
- O Não sei

O que usa no seu smartphone ou tablet? Escolha todas as opções que se aplicam

- ☐ Chamadas e SMS
- ☐ Acesso à Internet
- ☐ Aplicações originais do dispositivo (exemplos: bloco de notas, calculadora, etc.)
- ☐ Aplicações que instalou (exemplos: jogos, redes sociais, etc.)
- ☐ Outra: __________

Qual é o sistema operativo do seu dispositivo?

- O Android
- O iOS
- O Windows
- O Blackberry
- O Não sei
- O Outro: __________

II. Apontamentos

Costuma tomar nota de informações de saúde? O Sim O Não

De que tipo? Escolha todas as opções que se aplicam

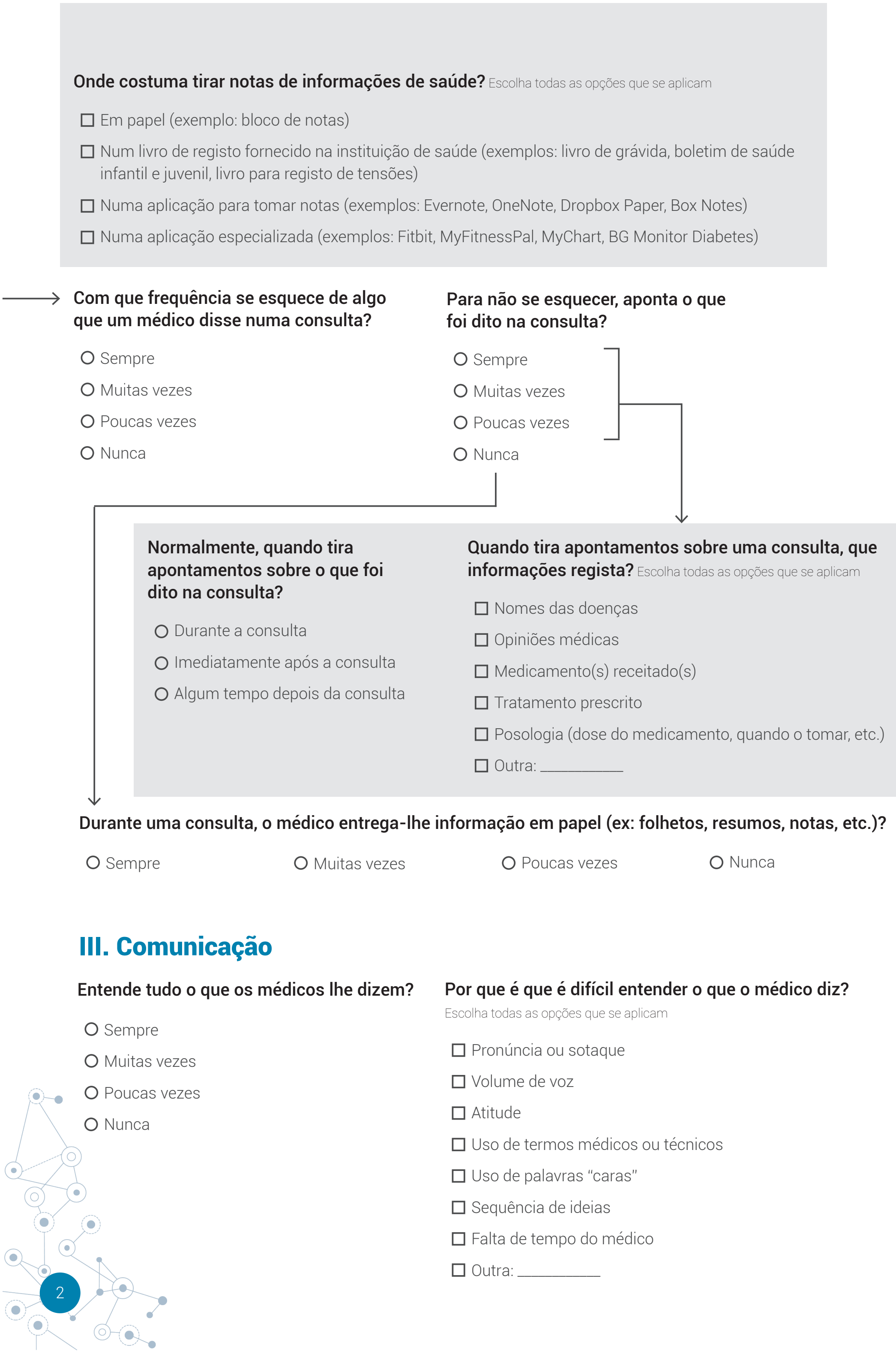
- ☐ Peso
- ☐ Calorias ingeridas e/ou despendidas
- ☐ Batimentos cardíacos ou pulso
- ☐ Tensão arterial
- ☐ Sintomas ou sinais clínicos
- ☐ Nome da(s) doença(s) que tem
- ☐ Pontos a discutir numa consulta futura
- ☐ Informações encontradas na Web
- ☐ Resultados de exames ou análises (exemplo: picar o dedo para saber o açúcar no sangue)
- ☐ Outro: __________

Sobre quem? Escolha todas as opções que se aplicam

- ☐ Sobre si próprio(a)
- ☐ Sobre um familiar
- ☐ Sobre um(a) amigo(a)
- ☐ Outro: __________

1

Figure A.1: First page of the questionnaire prototype.



Onde costuma tirar notas de informações de saúde? Escolha todas as opções que se aplicam

- ☐ Em papel (exemplo: bloco de notas)
- ☐ Num livro de registo fornecido na instituição de saúde (exemplos: livro de grávida, boletim de saúde infantil e juvenil, livro para registo de tensões)
- ☐ Numa aplicação para tomar notas (exemplos: Evernote, OneNote, Dropbox Paper, Box Notes)
- ☐ Numa aplicação especializada (exemplos: Fitbit, MyFitnessPal, MyChart, BG Monitor Diabetes)

Com que frequência se esquece de algo que um médico disse numa consulta?

- O Sempre
- O Muitas vezes
- O Poucas vezes
- O Nunca

Para não se esquecer, aponta o que foi dito na consulta?

- O Sempre
- O Muitas vezes
- O Poucas vezes
- O Nunca

Normalmente, quando tira apontamentos sobre o que foi dito na consulta?

- O Durante a consulta
- O Imediatamente após a consulta
- O Algum tempo depois da consulta

Quando tira apontamentos sobre uma consulta, que informações regista? Escolha todas as opções que se aplicam

- ☐ Nomes das doenças
- ☐ Opiniões médicas
- ☐ Medicamento(s) receitado(s)
- ☐ Tratamento prescrito
- ☐ Posologia (dose do medicamento, quando o tomar, etc.)
- ☐ Outra: ___________

Durante uma consulta, o médico entrega-lhe informação em papel (ex: folhetos, resumos, notas, etc.)?

O Sempre O Muitas vezes O Poucas vezes O Nunca

III. Comunicação

Entende tudo o que os médicos lhe dizem?

- O Sempre
- O Muitas vezes
- O Poucas vezes
- O Nunca

Por que é que é difícil entender o que o médico diz?
Escolha todas as opções que se aplicam

- ☐ Pronúncia ou sotaque
- ☐ Volume de voz
- ☐ Atitude
- ☐ Uso de termos médicos ou técnicos
- ☐ Uso de palavras "caras"
- ☐ Sequência de ideias
- ☐ Falta de tempo do médico
- ☐ Outra: ___________

2

Figure A.2: Second page of the questionnaire prototype.

Quando não percebe algo, pede ao médico para repetir ou reformular o que disse?

- ○ Sempre
- ○ Muitas vezes
- ○ Poucas vezes
- ○ Nunca

Quando não pede ao médico para repetir ou reformular o que disse, porque não o faz? Escolha todas as opções que se aplicam

- ☐ Falta de confiança
- ☐ Receio de parecer ignorante
- ☐ Não querer incomodar
- ☐ Alguém lhe explicará mais tarde
- ☐ Confia no médico e não precisa de saber mais
- ☐ Falta de tempo ou oportunidade
- ☐ Desinteresse
- ☐ Outra: ___________

IV. Aplicação

Numa aplicação para smartphones/tablets seria importante permitir...

	Não concordo	Indiferente	Concordo
Gravar a voz do seu médico durante a consulta e permitir-lhe ouvir essas gravações mais tarde.	○	○	○
Gravar a voz do seu médico durante a consulta e permitir-lhe consultar o texto do que foi dito nessas gravações mais tarde.	○	○	○
Fornecer definições dos termos médicos nos textos do que foi dito numa consulta.	○	○	○
Associar anotações às consultas e gerir outra informação de saúde.	○	○	○

Usaria uma aplicação que permitisse fazer o que foi indicado na pergunta anterior?

○ Sim ○ Talvez ○ Não

Se pudesse criar uma aplicação para resolver os temas aqui abordados, que outras utilidades gostaria que fossem implementadas? Como é que a aplicação funcionaria?

Obrigado!

3

Figure A.3: Third page of the questionnaire prototype.

Appendix B

Questionnaire

All of the questions included in the questionnaire and the answer options are listed in Figures B.1 to B.4 in their original Portuguese version.

HEALTH
TALKS

Este questionário foi desenvolvido no âmbito de uma dissertação de mestrado na Faculdade de Engenharia da Universidade do Porto e pretende aferir a eficácia da comunicação médico-paciente e dos métodos de gestão pessoal de informação de saúde. O inquérito demora aproximadamente 5 minutos a responder.

Por favor, assinale as suas respostas marcando com um 'X' os círculos ou quadrados que as antecedem; em algumas questões poder-lhe-á ser permitido dar mais do que uma resposta. Por vezes ser-lhe-á pedido no final de uma pergunta para saltar para outra pergunta especificada dependendo da resposta que tiver dado. Se nada for dito, prossiga para a questão seguinte.

A sua identidade permanecerá anónima.

Se quiser esclarecer alguma dúvida pode contactar o mestrando João Monteiro através do seguinte endereço de correio eletrónico: ei11055@fe.up.pt

Quando terminar, por favor coloque o questionário na caixa para o efeito que se encontra na receção.

Obrigado pela sua colaboração!

1. Dados demográficos

1.1. Género

- O Masculino
- O Feminino
- O Outro: ___________

1.2. Idade

- O 18-25 anos
- O 26-35 anos
- O 36-49 anos
- O 50-64 anos
- O 65+ anos

1.3. Habilitações (Selecione o grau de ensino mais elevado que tenha completado)

- O Não completei a 4.ª classe
- O 4.ª classe
- O 9.º ano
- O 12.º ano
- O Licenciatura ou bacharelato
- O Mestrado
- O Doutoramento

2. Apontamentos

2.1. Costuma tomar nota de informações de saúde?

- O Sim *(se esta for a sua resposta, continue para a pergunta **2.2.**)*
- O Não *(se esta for a sua resposta, salte para a pergunta **2.5.**)*

2.2. Tira notas de saúde sobre quem?

(Escolha todas as opções que se aplicam)

- ☐ Sobre si próprio(a)
- ☐ Sobre um familiar
- ☐ Sobre um(a) amigo(a)
- ☐ Outro: ___________

2.3. Que informações costuma apontar? (Escolha todas as opções que se aplicam)

- ☐ Peso
- ☐ Calorias ingeridas e/ou despendidas
- ☐ Batimentos cardíacos ou pulso
- ☐ Tensão arterial
- ☐ Sintomas ou sinais clínicos
- ☐ Nome da(s) doença(s) que tem
- ☐ Pontos a discutir numa consulta futura
- ☐ Informações encontradas na Web
- ☐ Resultados de exames ou análises (exemplo: picar o dedo para saber o açúcar no sangue)
- ☐ Outro: ___________

1

Figure B.1: First page of the written version of the questionnaires.

2.4. Onde costuma tirar notas de informações de saúde? (Escolha todas as opções que se aplicam)

- ☐ Em papel (exemplo: bloco de notas)
- ☐ Num livro de registo fornecido na instituição de saúde (exemplos: livro de grávida, boletim de saúde infantil e juvenil, livro para registo de tensões)
- ☐ Numa aplicação para tomar notas (exemplos: Evernote, OneNote, Dropbox Paper, Box Notes)
- ☐ Numa aplicação especializada (exemplos: Fitbit, MyFitnessPal, MyChart, BG Monitor Diabetes)

2.5. Com que frequência se esquece de algo que um médico disse numa consulta?

○ Sempre ○ Muitas vezes ○ Poucas vezes ○ Nunca

2.6. Para não se esquecer, aponta o que foi dito na consulta?

- ○ Sempre *(se esta for a sua resposta, continue para a pergunta* ***2.7.****)*
- ○ Muitas vezes *(se esta for a sua resposta, continue para a pergunta* ***2.7.****)*
- ○ Poucas vezes *(se esta for a sua resposta, continue para a pergunta* ***2.7.****)*
- ○ Nunca *(se esta for a sua resposta, salte para a pergunta* ***2.9.****)*

2.7. Normalmente, quando tira apontamentos sobre o que foi dito na consulta?

- ○ Durante a consulta
- ○ Imediatamente após a consulta
- ○ Algum tempo depois da consulta

2.8. Quando tira apontamentos sobre uma consulta, que informações regista? (Escolha todas as opções que se aplicam)

- ☐ Nomes das doenças
- ☐ Opiniões médicas
- ☐ Medicamento(s) receitado(s)
- ☐ Tratamento prescrito
- ☐ Posologia (dose do medicamento, quando o tomar, etc.)
- ☐ Outra: ___________

2.9. Durante uma consulta, o médico entrega-lhe informação em papel (exemplos: folhetos, resumos, notas)?

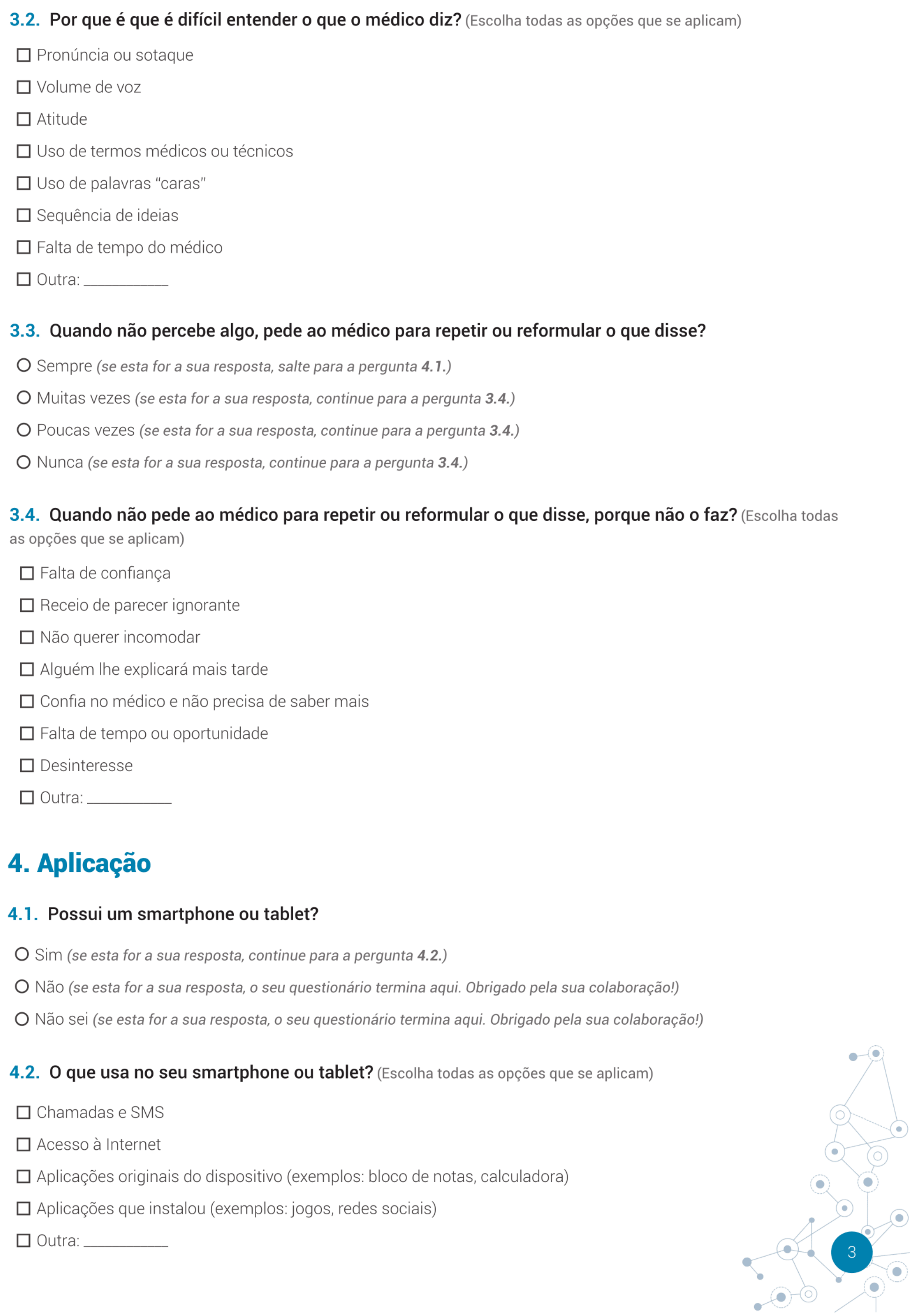
○ Sempre ○ Muitas vezes ○ Poucas vezes ○ Nunca

3. Comunicação

3.1. Entende tudo o que os médicos lhe dizem?

- ○ Sempre *(se esta for a sua resposta, salte para a pergunta* ***4.1.****)*
- ○ Muitas vezes *(se esta for a sua resposta, continue para a pergunta* ***3.2.****)*
- ○ Poucas vezes *(se esta for a sua resposta, continue para a pergunta* ***3.2.****)*
- ○ Nunca *(se esta for a sua resposta, continue para a pergunta* ***3.2.****)*

Figure B.2: Second page of the written version of the questionnaires.



3.2. Por que é que é difícil entender o que o médico diz? (Escolha todas as opções que se aplicam)

☐ Pronúncia ou sotaque

☐ Volume de voz

☐ Atitude

☐ Uso de termos médicos ou técnicos

☐ Uso de palavras "caras"

☐ Sequência de ideias

☐ Falta de tempo do médico

☐ Outra: ___________

3.3. Quando não percebe algo, pede ao médico para repetir ou reformular o que disse?

○ Sempre *(se esta for a sua resposta, salte para a pergunta **4.1.**)*

○ Muitas vezes *(se esta for a sua resposta, continue para a pergunta **3.4.**)*

○ Poucas vezes *(se esta for a sua resposta, continue para a pergunta **3.4.**)*

○ Nunca *(se esta for a sua resposta, continue para a pergunta **3.4.**)*

3.4. Quando não pede ao médico para repetir ou reformular o que disse, porque não o faz? (Escolha todas as opções que se aplicam)

☐ Falta de confiança

☐ Receio de parecer ignorante

☐ Não querer incomodar

☐ Alguém lhe explicará mais tarde

☐ Confia no médico e não precisa de saber mais

☐ Falta de tempo ou oportunidade

☐ Desinteresse

☐ Outra: ___________

4. Aplicação

4.1. Possui um smartphone ou tablet?

○ Sim *(se esta for a sua resposta, continue para a pergunta **4.2.**)*

○ Não *(se esta for a sua resposta, o seu questionário termina aqui. Obrigado pela sua colaboração!)*

○ Não sei *(se esta for a sua resposta, o seu questionário termina aqui. Obrigado pela sua colaboração!)*

4.2. O que usa no seu smartphone ou tablet? (Escolha todas as opções que se aplicam)

☐ Chamadas e SMS

☐ Acesso à Internet

☐ Aplicações originais do dispositivo (exemplos: bloco de notas, calculadora)

☐ Aplicações que instalou (exemplos: jogos, redes sociais)

☐ Outra: ___________

3

Figure B.3: Third page of the written version of the questionnaires.

4.3. Qual é o sistema operativo do seu dispositivo?

- O Android
- O iOS
- O Windows
- O Blackberry
- O Não sei
- O Outro: ___________

4.4. Numa aplicação para smartphones/tablets seria importante permitir...

	Não concordo	Indiferente	Concordo
Gravar a voz do seu médico, com a autorização deste, durante a consulta e permitir-lhe ouvir essas gravações mais tarde.	O	O	O
Gravar a voz do seu médico, com a autorização deste, durante a consulta e permitir-lhe consultar o texto do que foi dito nessas gravações mais tarde.	O	O	O
Fornecer definições dos termos médicos nos textos do que foi dito numa consulta.	O	O	O
Associar anotações às consultas e gerir outra informação de saúde.	O	O	O

4.5. Usaria uma aplicação que permitisse fazer o que foi indicado na pergunta anterior?

O Sim　　O Talvez　　O Não

4.6. Se pudesse criar uma aplicação para resolver os problemas descritos neste questionário, que outras utilidades gostaria que se acrescentassem? Como é que a aplicação funcionaria?

4

Obrigado!

Figure B.4: Fourth page of the written version of the questionnaires.

Questionnaire

Appendix C

HealthTalks' Interfaces

The following images show the most important interfaces for HealthTalks. The app's landing screen, the first contact with the user, can be seen in Figure C.1. When the user clicks on the button "Iniciar gravação", they are sent to the interface in Figure C.2, the recording screen. After it ends, an appointment is automatically created (Figure C.3) with the current date and the default patient name (though both can be changed). The appointment has another two tabs; the second (Figure C.4) relates to the recording and its transcription. There, the user can play back the audio of a consultation or transcribe it, as seen in Figure C.5. In the transcription, medical concepts are found by comparing the terms used with the ones present in the app's database. In Figure C.6, the user clicked on a definition and is given the option to discover more information about it, if they wish to do so. Figure C.7 shows the last tab associated with an appointment, in which a user can take notes of that appointment.

All of the appointments are listed in the interface shown in Figure C.8, which is accessible from the landing screen by clicking on "Consultas". Those appointments can be easily found through a search interface where the user inserts the terms they are looking for and where they expect to find them, as seen in Figure C.9. Finally, by clicking on "Ajuda e definições" in the landing page, the user gets redirected to the view shown in Figure C.10, in which they can consult explanatory texts, as well as change the default patient name, delete all consultations, and update the definitions database.

Figure C.1: HealthTalks' landing screen.

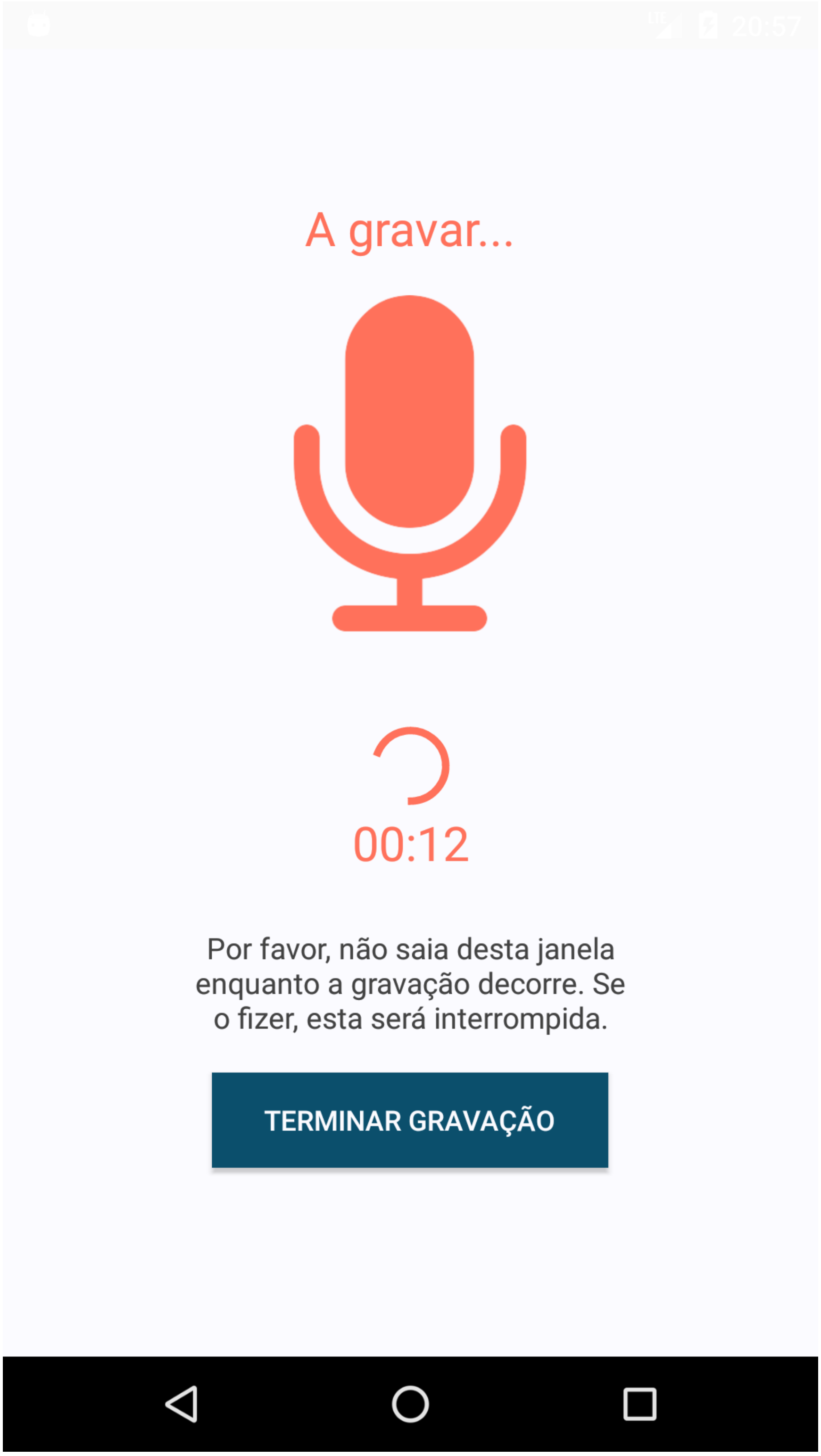


Figure C.2: Active screen when recording an appointment.

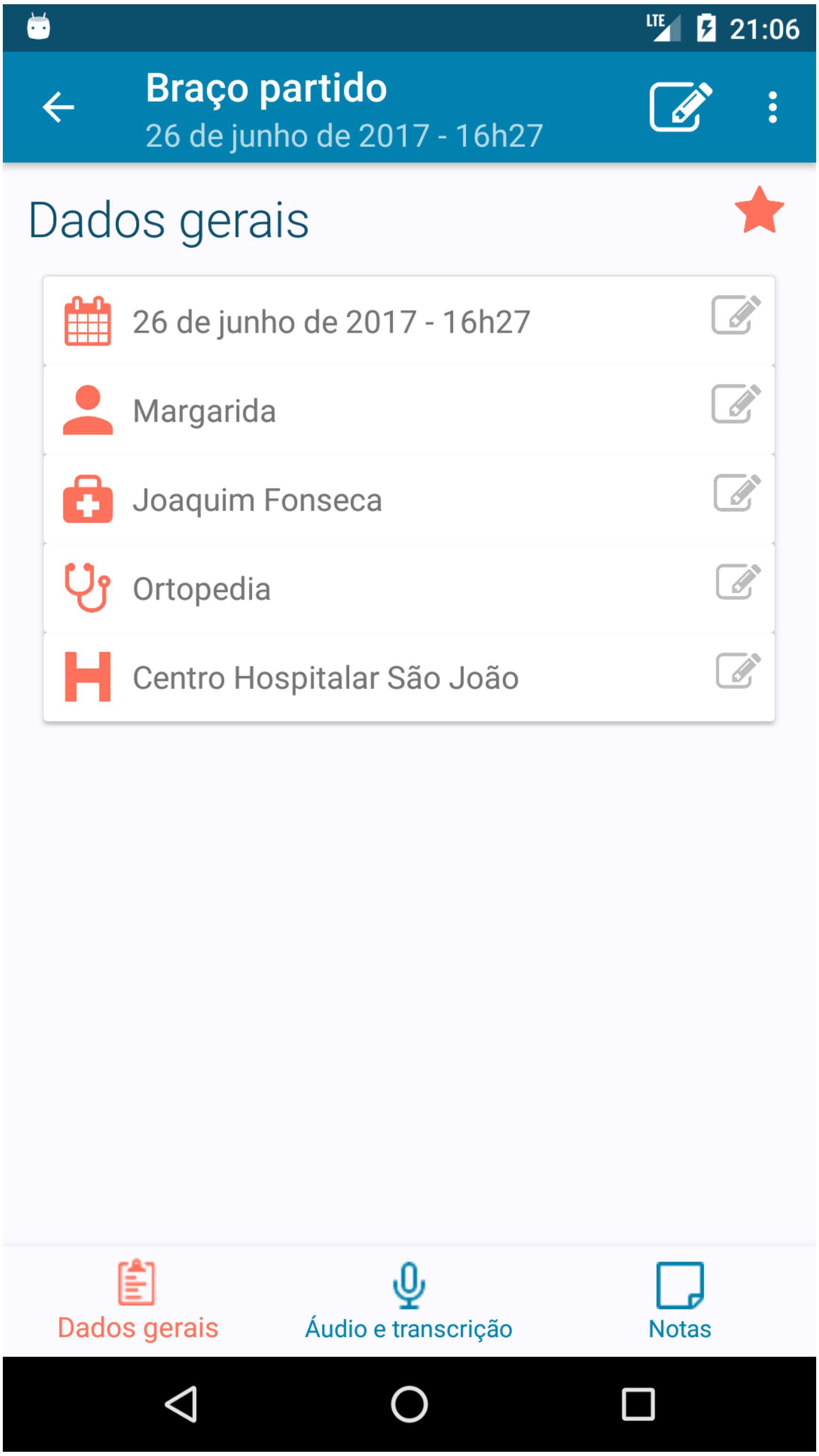


Figure C.3: General information about an appointment.

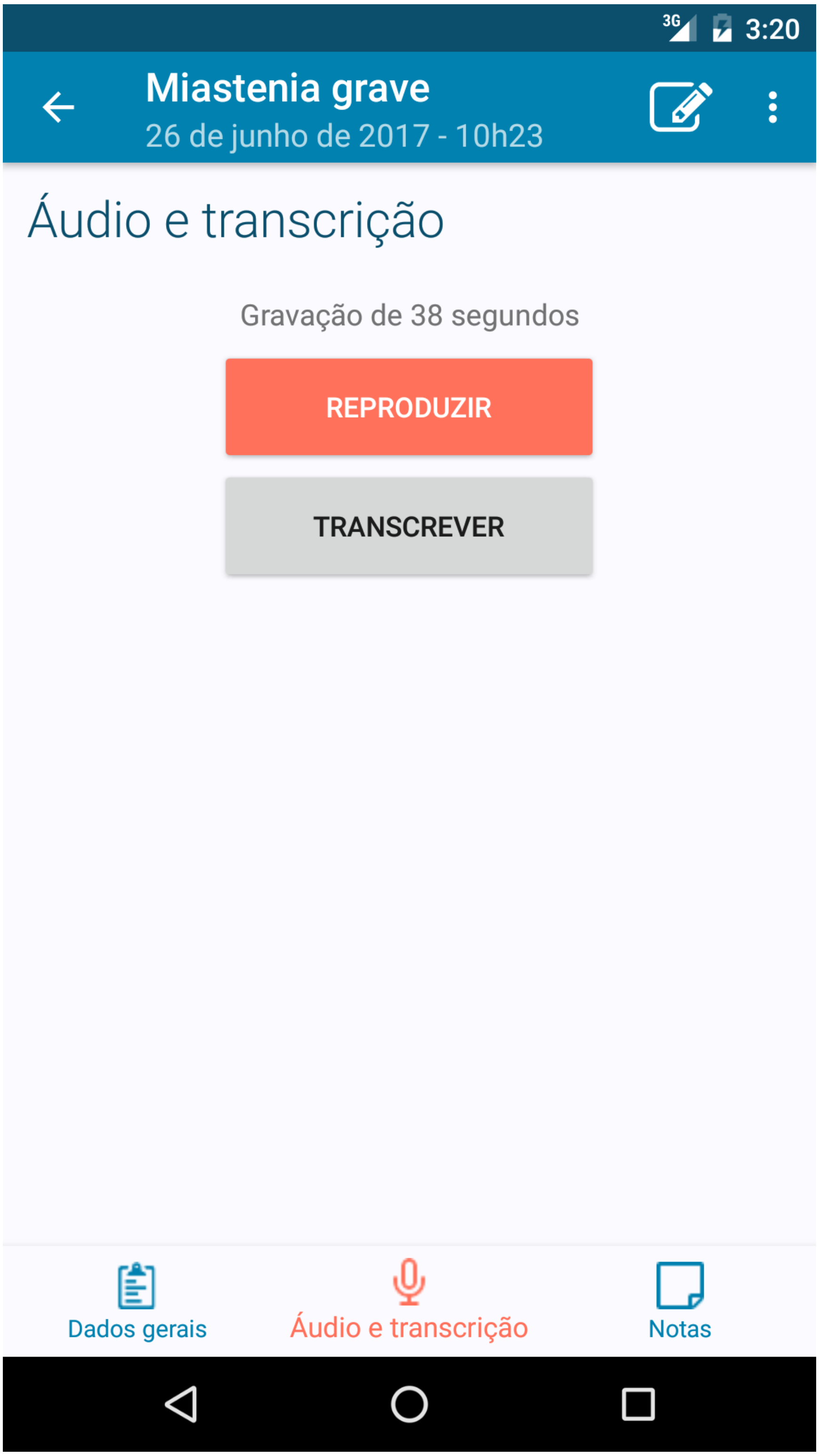


Figure C.4: Recording associated with an appointment.

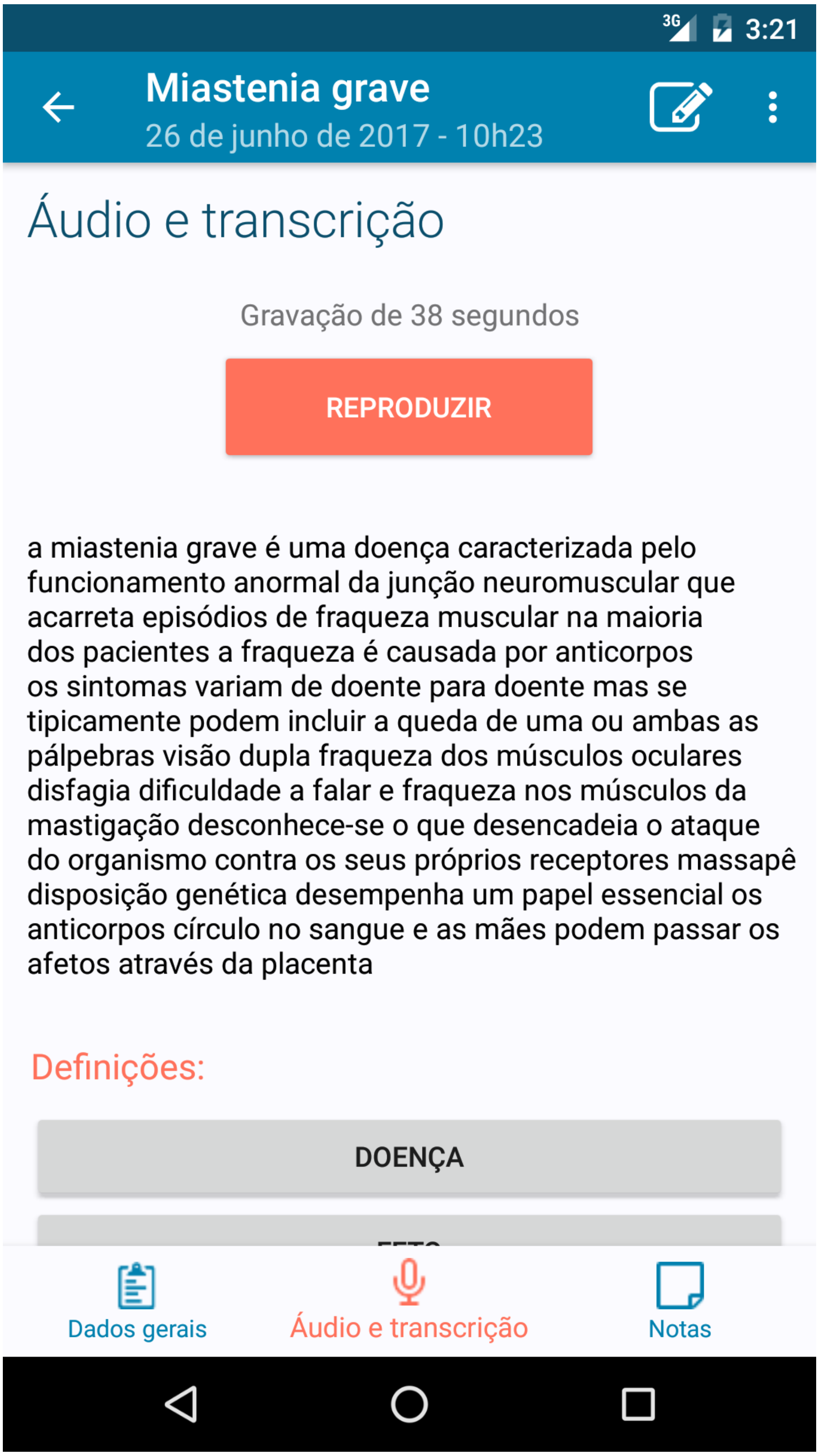


Figure C.5: Transcribed recording of an appointment.

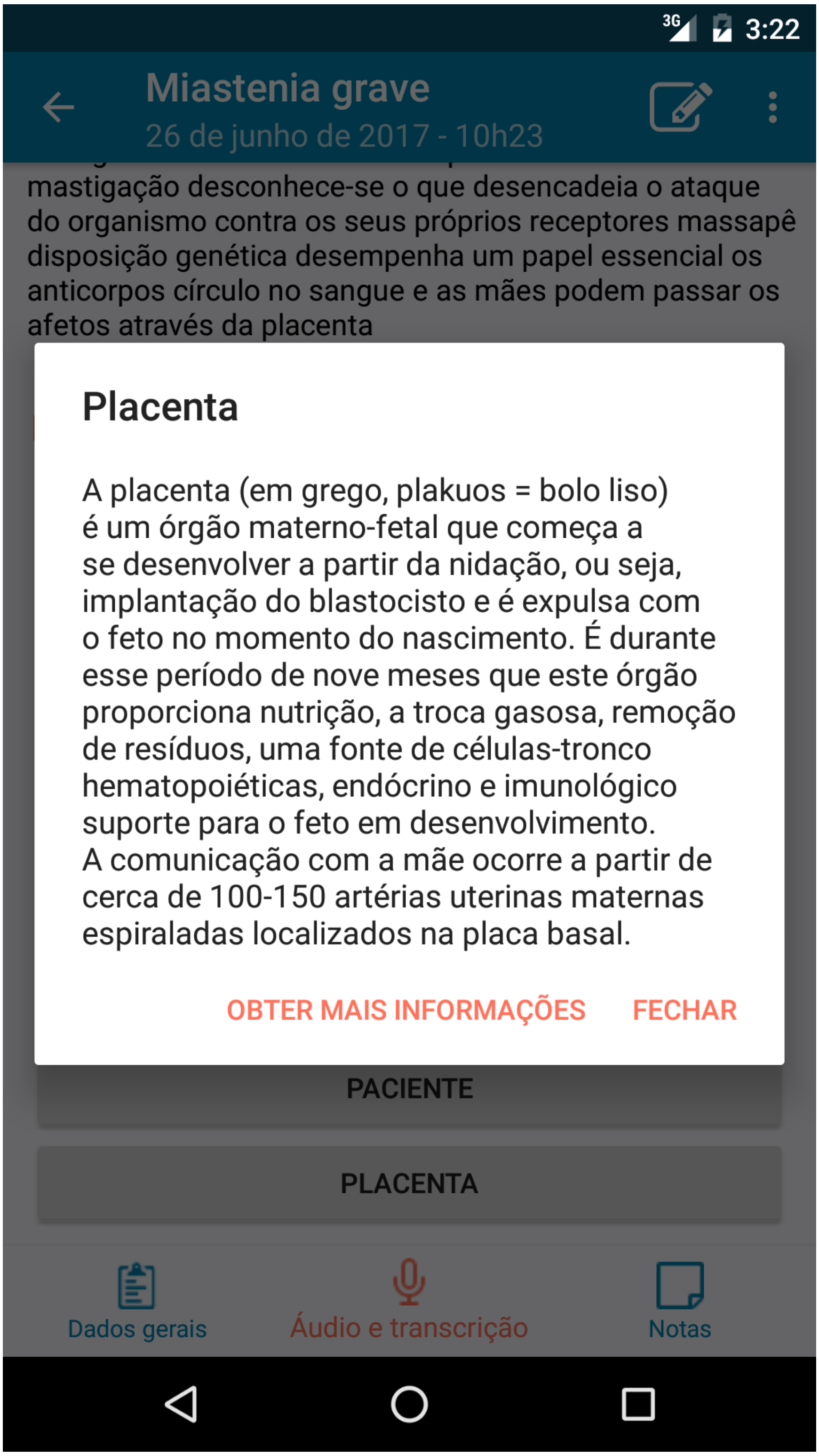


Figure C.6: Definition for one of the medical concepts found in the transcription.

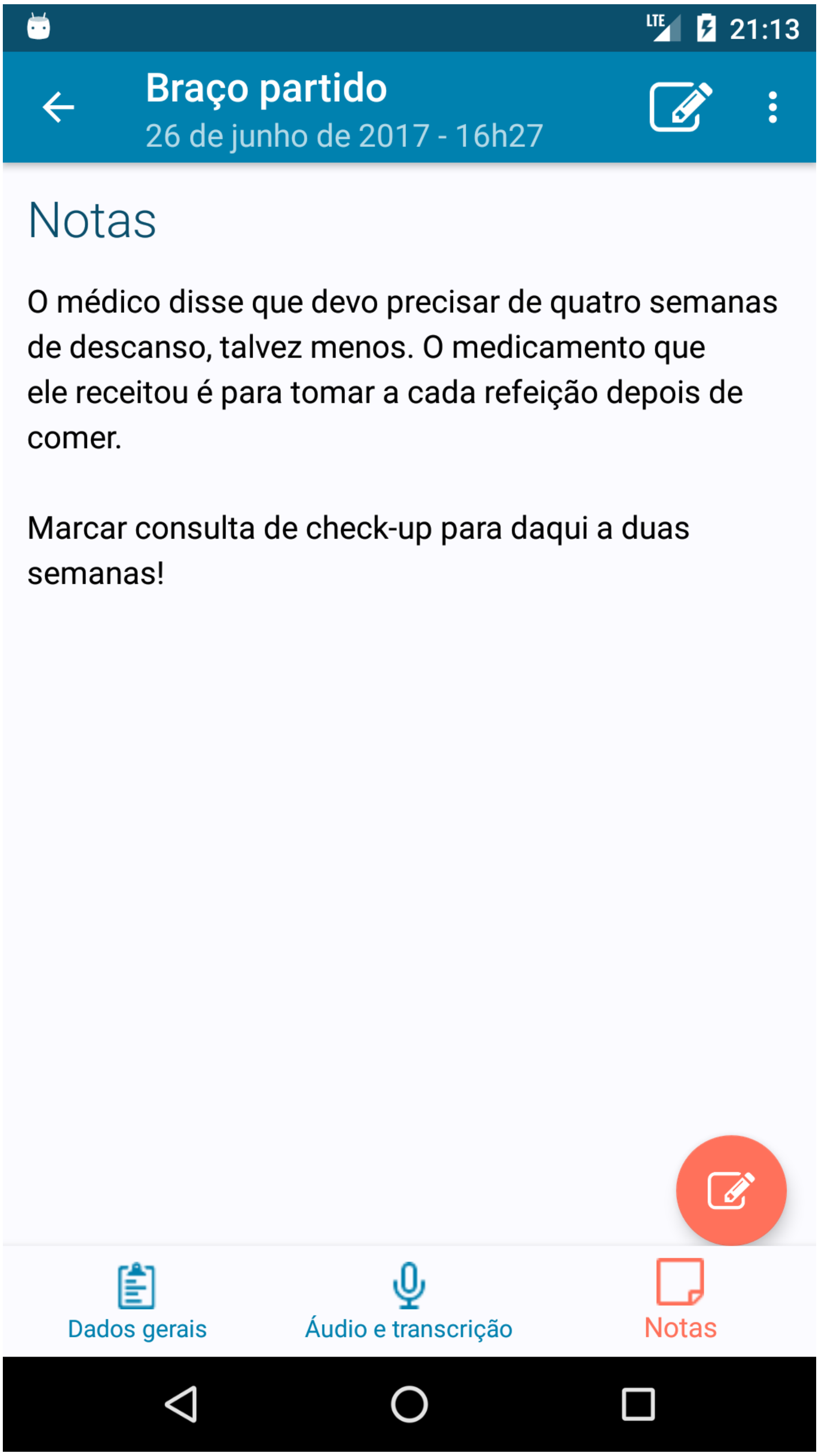


Figure C.7: Interface to take notes about an appointment.

Figure C.8: Interface for listing all the existing appointments.

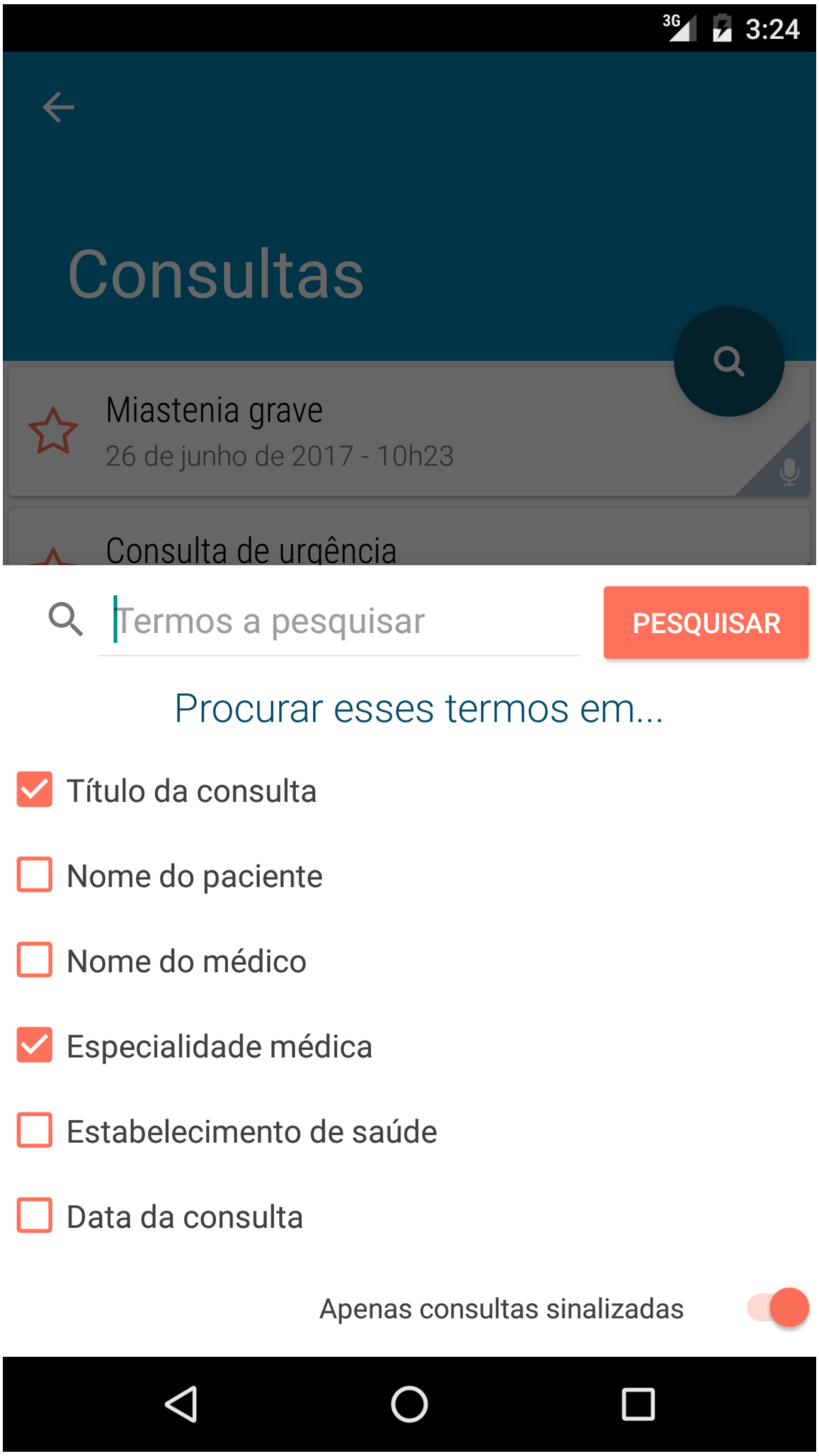


Figure C.9: Search interface.

Figure C.10: Interface presented after clicking on "Ajuda e definições".

HealthTalks’ Interfaces

Appendix D

Graphic Identity

There was an effort throughout the dissertation to create a coherent and recognisable look for HealthTalks in order to create a brand identity. That was achieved mostly by using specific colours, fonts, assets and guidelines in all the digital and printed media that permitted customised layouts, including the online and printed questionnaires, presentations, the documents delivered to the CHSJ's CES, the scientific poster presented in DSIE, the high-fidelity mockups and other assets like logos, banners, and even documents such as the consenting permit for UX testers.

HealthTalks' graphic identity started with the choice of the main colour palette. The first colour chosen was blue, since it is a primary colour and heavily associated with trust, peacefulness, truth, and wisdom. Many healthcare-related logos are blue because of that, and in that way this choice cemented HealthTalk's purpose as a healthcare-oriented solution. The "Bondi Blue" colour served as a basis of inspiration for the main blue chosen (R=0, G=129, B=175), since its use by Apple in the iMac G3 personal computer[1] marked the beginning of their notoriously divergent path in technological industrial design; this choice was an homage to their bold and successful design choices since that first milestone. This specific tone was then a mix between both of those established tones: the one mostly used in healthcare, and Apple's iMac G3 characteristic colour.

The selection of two more distinct shades of blue was a necessity so as to not overwhelm the canvas with the vibrant main blue. The process was very straightforward: two new secondary tones were created in order to accentuate the main blue, one considerably darker (R=11, G=79, B=108) and another lighter (R=0, G=171, B=231). A fourth colour was born out of necessity for situations where the lighter version was deemed too bright, and hence a grayish-blue hue (R=167, G=188, B=205) was chosen to use in small details.

The accent colour chosen was orange, since it can give emphasis to highlighted content and complements blue while transmitting energetic, stimulating and strong feelings. Its use was purposefully limited for that exact reason: while it balances the soothing effects of the blue colour, if used in excess it may transpire aggressiveness. It was hence used sparingly, mostly to emphasise

[1] A computer Apple itself considers an "unwavering commitment to craft the ultimate desktop computer", as seen in https://www.apple.com/imac/then-and-now/

noteworthy elements. The particular shade picked (R=255, G=113, B=91), a slightly darker variation of Crayola's "Atomic Tangerine", was selected because it evokes a call to action and does not clash with the lighter and darker blues chosen previously, so it can be used against them in the foreground.

Throughout the development there was a conscious effort to avoid the use of pure black and white to avoid harsh contrasts. It was decided to use an extremely dark blue (R=6, G=24, B=33) as a black replacement, and an off-white tone (R=251, G=251, B=255) as a white substitute. The exception for this last colour's use was when designing documents to be printed, since it would make them unnecessarily stylised.

One of the elements used more often throughout the different layouts was a mesh of randomly connected dots with different sizes and line styles. That vector was used mostly in corners, and always in a subtle contrast with its background so as to not be thought of as an important element. Its structure is reminiscent of graph data structures, with vertices and edges, which relates to the use of technology in the project. At the same time, those dots and lines are also similar to neurons in a human brain, with the lines standing for their axons. This can be seen as a nod to the educational and information management components of the project.

The fonts used were "Roboto" and its variants, mainly "Roboto Light", "Roboto Black", and "Roboto Condensed". This choice laid mostly with the fact that Android was the selected platform for the app, and this is its default font, as well as the font used in several other Google services. It has been described by Google as a "modern, yet approachable" and "emotional" font, with a geometric frame that contrasts with its relaxed curves [OBr11]. The usage of the different weights had mostly to do with the specific needs of each instance, with the default choice being "Roboto". The light version was preferred for small to medium strings with a large font size; condensed for strings that could grow to be significantly larger than predicted; and black for small titles and highlights.

A sample of colours and fonts can be seen in Figure D.1. The design language used throughout all media was flat design with occasional drop shadows.

Figure D.1: Colour and font samples; example of the dot mesh on both backgrounds.

Appendix E

Declaration of Consent

What follows in Figure E.1 is the consent form signed by all of the usability testers prior to participating in the test.

DECLARAÇÃO DE **CONSENTIMENTO**

Designação do Estudo:
Testes de usabilidade da aplicação HealthTalks

Eu, abaixo-assinado, (nome completo do(a) participante)

____,

realizei testes de usabilidade da aplicação "HealthTalks", parte constituinte da dissertação de mestrado do aluno João Monteiro, número 201104369, intitulada "HealthTalks - a mobile app to recognize medical speech", no âmbito do Mestrado Integrado em Engenharia Informática e Computação da Faculdade de Engenharia da Universidade do Porto. Foi-me comunicado que os dados recolhidos neste trabalho académico (respostas aos questionários, gravação áudio da entrevista, captura de ecrã, registo fotográfico, e dados pessoais fornecidos) poderão ser utilizados no estudo de investigação em epígrafe, para o qual é pedida a minha participação. Foi-me garantido que as informações recolhidas nunca serão publicamente associadas ao meu nome. Foi-me dada oportunidade de fazer as perguntas que julguei necessárias, e para todas obtive resposta satisfatória. Foi-me afirmado que tenho o direito de decidir livremente aceitar ou recusar a minha participação no estudo. Foi-me dado todo o tempo de que necessitei para reflectir sobre esta proposta de participação. Nestas circunstâncias, decido livremente aceitar participar neste projecto de investigação, tal como me foi apresentado pelo investigador.

Data: 22 de maio de 2017

Assinatura do(a) participante: ____

Assinatura do investigador: ____

(João Monteiro)

Figure E.1: Declaration of consent for usability testers.

Appendix F

Technology Readiness Index

As mentioned in Section 8.2.2, the TRI 2.0 scale proposed by Parasuraman and Colby [PC15] was used to characterise the UX testers according to their acceptance of new technology. Since the original questions were in English and no Portuguese version was available, we translated them to Portuguese. That translation did not follow scientific norms and procedures for the translation and adaptation of research instruments such as the ones established by the World Health Organization (WHO) [Wor10], since such a process would not be feasible for this study. However, there was great care in guaranteeing the closest possible translation, in order not to change the instrument, while at the same time making the language fluid and clear, just like it is in the original version. In Table F.1 we can see each of the 16 items of TRI 2.0 as well as the respective translations.

Table F.1: TRI 2.0's 16 item scale and developed translation. The three letters in each code represent the category that statement belongs to: OPT for optimism, INN for innovativeness, DIS for discomfort, and INS for insecurity.

Code	Original	Translation
OPT1	New technologies contribute to a better quality of life	Novas tecnologias contribuem para uma melhor qualidade de vida
OPT2	Technology gives me more freedom of mobility	A tecnologia dá-me mais liberdade de mobilidade
OPT3	Technology gives people more control over their daily lives	A tecnologia dá às pessoas mais controlo sobre a sua vida diária
OPT4	Technology makes me more productive in my personal life	A tecnologia torna-me mais produtivo(a) na minha vida pessoal
INN1	Other people come to me for advice on new technologies	Outras pessoas vêm ter comigo para conselhos sobre novas tecnologias
INN2	In general, I am among the first in my circle of friends to acquire new technology when it appears	Em geral, sou um(a) dos(as) primeiros(as) do meu grupo de amigos a adquirir nova tecnologia quando ela surge

INN3	I can usually figure out new high-tech products and services without help from others	Normalmente eu consigo-me desenvencilhar com produtos e serviços novos de alta tecnologia sem a ajuda de outras pessoas
INN4	I keep up with the latest technological developments in my areas of interest	Eu mantenho-me a par dos desenvolvimentos tecnológicos mais recentes nas minhas áreas de interesse
DIS1	When I get technical support from a provider of a high-tech product or service, I sometimes feel as if I am being taken advantage of by someone who knows more than I do	Quando recebo apoio técnico de um fornecedor de um produto ou serviço de alta tecnologia, por vezes sinto que alguém que sabe mais do que eu está a tentar aproveitar-se de mim
DIS2	Technical support lines are not helpful because they don't explain things in terms I understand	As linhas de apoio técnico não são prestáveis porque não explicam as coisas em termos que eu compreenda
DIS3	Sometimes, I think that technology systems are not designed for use by ordinary people	Às vezes penso que os sistemas tecnológicos não foram feitos para serem usados por pessoas comuns
DIS4	There is no such thing as a manual for a high-tech product or service that's written in plain language	Não há nenhum manual de um produto ou serviço de alta tecnologia que esteja escrito em linguagem simples
INS1	People are too dependent on technology to do things for them	As pessoas dependem demasiado em tecnologia que faz as coisas por elas
INS2	Too much technology distracts people to a point that is harmful	Demasiada tecnologia distrai as pessoas a ponto de ser nocivo
INS3	Technology lowers the quality of relationships by reducing personal interaction	A tecnologia diminui a qualidade das relações ao reduzir a interação pessoal
INS4	I do not feel confident doing business with a place that can only be reached online	Eu não sinto confiança a fazer negócios com um local que só pode ser contactado online